French predicates selecting the subjunctive mood under the microscope: the emotive factor

Baunaz, L

DOI: https://doi.org/10.1075/rllt.11

Romance Languages and Linguistic Theory II

Romance Languages and Linguistic Theory II

Selected papers from the 44th Linguistic Symposium on Romance Languages (LSRL), London, Ontario

edited by
Silvia Perpiñán
David Heap
Itziri Moreno-Villamar
Adriana Soto-Corominas

John Benjamins Publishing Company

Romance Languages and Linguistic Theory 11

Romance Languages and Linguistic Theory (RLLT)

ISSN 1574-552X

The yearly events of both 'Going Romance' and 'Linguistic Symposium on Romance languages' feature research in formal linguistics of Romance languages, in the domains of syntax, morphology, phonology and semantics. Each volume brings together a peer-reviewed selection of papers that were presented at one of the meetings, aiming to provide a representation of the spread of topics at that conference, and of the variety of research carried out nowadays on Romance languages within theoretical linguistics.

For an overview of all books published in this series, please see
http://benjamins.com/catalog/rllt

Editor

Frank Drijkoningen
Utrecht University

Volume 11

Romance Languages and Linguistic Theory 11. Selected papers from
the 44th Linguistic Symposium on Romance Languages (LSRL), London, Ontario
Edited by Silvia Perpiñán, David Heap, Itziri Moreno-Villamar
and Adriana Soto-Corominas

Romance Languages and Linguistic Theory 11

Selected papers from the 44th Linguistic Symposium on Romance Languages (LSRL), London, Ontario

Edited by

Silvia Perpiñán
David Heap
Itziri Moreno-Villamar
Adriana Soto-Corominas
University of Western Ontario

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/rllt.11
Cataloging-in-Publication Data available from Library of Congress:
LCCN 2017015072 (PRINT) / 2017037562 (E-BOOK)

ISBN 978 90 272 0391 5 (HB)
ISBN 978 90 272 6534 0 (E-BOOK)

John Benjamins Publishing Company · https://benjamins.com

Table of contents

Introduction 1
Silvia Perpiñán, David Heap, Itziri Moreno-Villamar and Adriana Soto-Corominas

Part I. Syntax-semantics

CHAPTER 1
Embedding verbs and subjunctive mood: The emotive factor 9
Lena Baunaz

CHAPTER 2
Towards a unified treatment of Spanish copulas 33
María J. Arche, Antonio Fábregas and Rafael Marín

CHAPTER 3
How French sheds new light on scalar particles 53
Isabelle Charnavel

CHAPTER 4
Pluralities of events: Semelfactives and a case of 'single event' nominalisation 77
Marta Donazzan and Lucia M. Tovena

Part II. Morphosyntax

CHAPTER 5
Laísmo and "le-for-les": To agree or not to agree 101
Adolfo Ausín and Francisco J. Fernández-Rubiera

CHAPTER 6
The morphological markedness of φ: Evidence from perfective auxiliaries in Southern Italian dialects 127
Giuseppe Torcolacci

CHAPTER 7
Partial subject paradigms and feature geometry in Northern Occitan dialects 147
Michèle Oliviéri, Jean-Pierre Lai and David Heap

CHAPTER 8
Automatic detection of syntactic patterns from texts with application to Spanish clitic doubling 169
Bruno Estigarribia

Part III. Bilingualism and language acquisition

CHAPTER 9
Voice quality transfer in the production of Spanish heritage speakers and English L2 learners of Spanish 191
Ji Young Kim

CHAPTER 10
Null subjects in the early acquisition of English by child heritage speakers of Spanish 209
Jennifer Austin, Liliana Sánchez and Silvia Perez-Cortes

CHAPTER 11
Return to Frenchville: Tracing a near-merger from legacy data 229
Barbara E. Bullock and Jenna Nichols

CHAPTER 12
The processing of intrasentential anaphoric subject pronouns in L2 Spanish 247
Juan P. Comínguez, Nuria Sagarra, Aurora Bel and Estela García-Alcaraz

Index 265

Introduction

Silvia Perpiñán, David Heap, Itziri Moreno-Villamar
and Adriana Soto-Corominas
University of Western Ontario

On May 2–4, 2014, The University of Western Ontario and in particular the Faculty of Arts & Humanities, the Department of French Studies and the Department of Modern Languages & Literatures had the privilege of organizing the 44th Linguistic Symposium on Romance Languages in London Ontario, Canada. The symposium had over 60 talks and posters, including the keynote presentations by our invited speakers: Joyce Bruhn de Garavito, from the hosting university, Maria Teresa Guasti, from the Università di Milano-Bicocca, José Ignacio Hualde, from the University of Illinois, Urbana-Champaign, and Viviane Déprez, from Rutgers University. We believe that we achieved our initial purpose of having different Romance languages and varieties, as well as different linguistic sub-disciplines well represented.

The Symposium was accompanied by a pre-conference event on May 1st, 2014, at Museum London. This pre-conference event included the exhibit 'French in Canada' from the Canadian Language Museum, and two community-oriented talks, one in French by François Poiré "Contributions des études sur la communauté franco-ontarienne à la recherche en linguistique", and in Spanish by Joyce Bruhn de Garavito, "La importancia de nuestro idioma en Canadá: hablemos español con la familia".

The conference itself was a total success: everything went smoothly, and most importantly, the quality of the talks and poster presentations was outstanding. Of 66 presentations at the conference, 28 were submitted for publication, a number which we had to reduce to just twelve exceptional articles. The selection of the papers you have in this volume has been a difficult task, as we had a rigorous page-limit for the edited volume. We are very happy that the publication of the LSRL proceedings has moved, beginning with the last volume, LSRL 43, to the 'Romance Languages and Linguistic Theory' series within John Benjamins, so that volume editors no longer have a topic restriction for theses volumes. This series was already publishing the proceedings of the European 'Going Romance', so these two sister conferences now have a meeting point in the same publication series.

DOI 10.1075/rllt.11.int

We have divided the present volume into three major representative subfields, with four articles each: Syntax-Semantics, Morphosyntax, and Bilingualism and Language Acquisition.

In the first article of the Syntax-Semantics section, "Embedding Verbs and the Subjunctive Mood: the Emotive Factor", Lena Baunaz explores the uses of the subjunctive mood in French, in particular systematic grammatical differences in five classes of verbs which govern dependent clauses. She uses a nanosyntactic analysis to show that *veridicality* cannot be the deciding factor in mood selection. The claim here is that at the submorphemic level there is an *emotive* feature of certain predicates which requires the subjunctive mood. This feature leads to a proposed unified account of predicates that take the subjunctive, including impersonal structures.

The second article in this section "Towards a Unified Treatment of Spanish Copulas", by Arche, Fábregas, and Marín adds to the ongoing debate of the distribution of the Spanish copulas *ser* and *estar* by accounting for their alternation in adjectival and passive clauses in a unified way. They propose that the properties of passive clauses are due to the properties of the copulas and not the participles, and further argue that only *estar* has an additional component of central coincidence with a stative nature.

Charnavel analyzes French scalar particles *même*, *quand même*, *ne serait-ce que,* and *seulement* and compares them to English particles *even* and *only.* Her article, "How French Sheds New Light on Scalar Particles" proposes a new theory based on specific characteristics such as scalarity, additivity, and exclusivity and provides new empirical evidence about these French particles, which, despite widespread assumptions, behave differently from their English counterparts.

The last article of the Syntax-Semantics section is by Donazzan and Tovena and investigates the semantics of semelfactive predicates in Italian. In "Pluralities of Events: Semelfactives and a Case of 'Single Event' Nominalisation", the authors analyze the notion of plurality and unity of events by looking at the two possible readings that *ata*-nominalisations, i.e.: *nuotata*, *ombrellata,* receive in instrument semelfactive verbs. They conclude that semelfactives, in their processive readings, have to considered activity predicates.

The second section of the volume includes articles on Morphosyntax, beginning with "Laísmo and 'le-for-les': To Agree or not to Agree?", by Adolfo Ausín and Francisco J. Fernández-Rubiera, a novel exploration of an old problem in nonstandard Spanish pronominal paradigms. They propose a unified account for three apparently unrelated phenomena: the presence of the accusative clitic, the presence/absence of number agreement in the dative clitic ('le-for-les'), and the presence/absence of gender agreement in the dative clitic in laísta dialects ('le-for-la'). They propose that agreement in these two nonstandard dative clitic constructions is related

to the required presence of accusative clitics; at the same time, they assume that dative clitics are composed of two different morphemes: an applicative morpheme and optionally an agreement one. The presence of agreement in dative clitics occurs when the applicative morpheme is combined with the agreement morpheme; lack of agreement is obtained when only the applicative morpheme surfaces.

In this section's second article, Giuseppe Torcolacci examines featural composition of verbal morphology with "The Morphological Markedness of φ. Evidence from Perfective Auxiliaries in Southern Italian Dialects". Using the Distributed Morphology framework, he posits that in these dialects, Spell-Out checks the φ features of perfect auxiliaries to see if their degree of markedness is the same as that of their Tense features, and only then allows these features to be overtly marked, a process characterized as Default Marking.

The third article in this section, "Partial Subject Paradigms and Feature Geometry in Northern Occitan dialects" by Michèle Oliviéri, Jean-Pierre Lai, and David Heap, considers the status of subject pronouns in transitional dialects which are situated between grammars which allow null-subjects and grammars that require an overt subject for all finite verbs. The paradigms in this transition area are structured differently from those in better-known Northern Italian dialects, due to the presence of a generic pronoun used for meteorological and impersonal constructions.

"Automatic Detection of Syntactic Patterns from Texts with Application to Spanish Clitic Doubling," by Bruno Estigarribia, is a corpus-based study that proposes algorithms for finding examples of direct object clitic doubling in samples of Spanish texts, particularly from the Web. Using a part-of-speech tagger and java code, this procedure matches doubled NPs with direct object clitics, between 50% and 100% of the time depending on the types of text examined. This approach shows the value of Natural Language Processing at the intersection of computational and corpus linguistics.

The third section, Bilingualism and Language Acquisition, begins with a paper on "Voice Quality Transfer in the Production of Spanish Heritage Speakers and English L2 Learners of Spanish" by Ji Young Kim. This study explores the use of creaky voice by U.S. heritage speakers of Spanish, and English L2 learners of Spanish as compared to Mexican Spanish monolinguals. Using an experimental reading task, and calculating the participants' relative amplitude of the first two harmonics in utterance-final positions, Kim found that bilingual speakers, but not monolinguals speakers (creaky voice is not a feature in Mexican Spanish), transferred creaky voice from American English into Spanish, particularly female speakers. This result shows that voice quality, like other linguistic features, can be transferred from one language to another.

"Null Subjects in the Early Acquisition of English by Child Heritage Speakers of Spanish," by Jennifer Austin, Liliana Sánchez, and Silvia Perez-Cortes looks at acquisition data in a non-Romance language in a bilingual context. Using two oral production tasks and detailed categories for coding discourse functions, subject productions were analyzed regarding their type (overt or null) and pragmatic function. The authors found that heritage Spanish-English bilinguals produced a higher rate of null subjects in English compared to age-matched monolinguals, which is consistent with the Multiple Grammars hypothesis, namely that bilinguals have available two competing grammars of English, one which cannot license null subjects, and a second one which treats English as a Spanish-like language, licensing null subjects.

In their acoustic study "Return to Frenchville: Tracing a Near-merger from Legacy Data", Barbara E. Bullock and Jenna Nichols look at the convergence of front mid rounded vowels /ø, œ/ with a rhoticized schwa [ɚ] in the French of two of the last speakers of a legacy variety spoken in Pennsylvania. While very similar in the previous generation, a comparison of acoustic data shows that these vowels in fact differed in duration and spectral quality. This near-merger in the second-last generation seems to have made possible a demerger due to rhotacization in the last generation, a pattern which is similar to the ones found in other varieties of French that do not suffer from the same degree of isolation.

The concluding paper of this section and of the volume, "The Processing of Intrasentential Anaphoric Subject Pronouns in L2 Spanish," by Juan Pablo Comínguez, Nuria Sagarra, Aurora Bel, and Estela García-Alcaraz, is an experimental study of Spanish learners' ability to process intra-sentential interpretive patterns of null and overt pronouns in native and non-native Spanish, and the extent to which those patterns are modulated by L1 transfer. With online and offline techniques, they tested English and Arabic advanced learners of Spanish, and found that both groups had a strong subject bias in backwards anaphoric interpretation, displaying convergence with the native preferences, regardless of the L1 preference.

This conference would not have been possible without generous support from a number of sources. For funding, we would like to thank the Social Sciences and Humanities Research Council (Connection Grant 611–2013–0054), the Faculty of Arts and Humanities, Research Western, the Department of Modern Languages and Literatures, the Department of French Studies, the Linguistics Program – particularly we would like to thank Rob Stainton – the Department of Philosophy, the Faculty of Arts and Humanities, the Faculty of Social Science, and the Faculty of Information and Media Studies. Our colleague Ileana Paul from the Department of French Studies was a key part of our team in developing, organizing and hosting LSRL44, and we are grateful for her help and guidance. The many peer reviewers

who evaluated both the abstracts for the conference and the papers for this volume deserve our heart-felt thanks for their anonymous contributions. Finally, and just as importantly, the many UWO student volunteers played a vital role in making this conference a success for us and for our colleagues who we hosted at the University of Western Ontario.

PART I

Syntax-semantics

CHAPTER 1

Embedding verbs and subjunctive mood
The emotive factor

Lena Baunaz
Ghent University & University of Zürich

I give a uniform account of French mood from a semantic and nano-syntactic perspective by revising the criteria for the distribution of indicative vs. subjunctive mood. I claim that the subjunctive is selected by the semantic property of *emotivity* encoded in the main clause verb. I develop a definition and a syntactic representation of emotive vs. non-emotive verbs and a set of diagnostics for the two classes. I also reexamine the notion of veridicality and show that it is not the determining factor in the choice of mood (contra Giannakidou 1998, 2009, a.o.).

Keywords: mood, dependent clauses, French, emotivity, nano-syntax, lexical semantics

1. Introduction

There is little agreement as to what constitutes the subjunctive cross- and intra-linguistically. It is nevertheless generally accepted that it is a *dependent mood* selected under verbs with special features (Quer 2009; Giannakidou 1998, 2009, a.o). On the basis of Modern Greek (MG), Giannakidou proposes that the crucial semantic factor regulating mood choice is (non)-veridicality. If a propositional attitude verb triggers at least one truth inference over its complement, it is veridical and triggers the indicative, (1a). If not, it is non-veridical, and triggers the subjunctive, (1b).

(1) a. *Nomizo* ***oti*** *efije o Janos* (MG)
I think IND left3.SG the John
'I think that John left.'

b. *Thelo* ***na*** *fiji o Janos* (Socanac 2011:51, (4))
I want SUBJ leave the John
'I want John to leave.'

DOI 10.1075/rllt.11.01bau

According to (2), an embedded proposition has to be true for at least one individual (the Subject of the main verb and/or the Speaker), in all the worlds of a relevant model:

(2) **Veridicality** (Giannakidou 1998: 1889)
a propositional operator *F* is veridical iff from the truth of *Fp* we can infer that *p* is true according to some individual *x* (i.e. in some individual *x*'s epistemic model)

With *factives* (*regret*, *know, realize*), both Speakers and Subjects are committed to the truth of the embedded proposition: they are strongly veridical; with *believe*, only the Subject is committed to the truth of the embedded proposition, it is veridical.

The mood-veridicality correlation runs into empirical problems when applied to Romance (Quer 2009) because unlike MG, Romance emotive factive predicates trigger the subjunctive mood, (3); these verbs presuppose the truth of their complements (Kiparsky & Kiparsky 1971). Under (2), *regret* should select the indicative, contrary to facts.

(3) *Roger regrette que Stan ait perdu*
R. regrets that S. has.SUBJ lost
I investigate the syntax and semantics of the verbs in (4):

(4) a. **Verbs of saying**: *dire* 'say', *observer* 'observe'…
b. **Epistemic**: *réaliser* 'realize', *se rappeler* 'remember'…
c. **Emotive factives**: *regretter* 'regret'…
d. **Desire**: *préférer* 'prefer', *souhaiter* 'wish', *vouloir* 'want'…
e. **Directive**: *suggérer* 'suggest', *insister* 'insist'…

These verbs coincide with systematic grammatical differences. Trivially the only difference between (5) and (6) is the morphological mood suffixed onto the embedded verb (subjunctive *parte* vs. indicative *part*). Verbs of saying and epistemic verbs take the indicative, (5), emotive factive, desire and directive verbs, the subjunctive, (6). Also, Subjects embedded under verbs selecting subjunctive CPs show obligatory obviation, (5), while embedded Subjects under verbs selecting indicative CPs do not (Ruwet 1984; Costantini 2009, a.o.), (6).

(5) $Roger_{i}$ ***{insiste/suggère}/regrette/veut*** $qu'il_{*i/j}$ *parte*
R. {insists/suggests}/regrets/wants that he leave.SUBJ

(6) $Roger_{i}$ ***réalise/dit*** $qu'il_{i/j}$ *part*
R. realizes/dit that he leaves.IND

Finally, the matrix predicates in (8) allow agent-oriented modification; those in (7) do not:

(7) *$Roger_i$ ***{insiste/suggère}/regrette/veut*** *intelligemment qu'$il_{*i/j}$* *parte*
R. {insists/suggests}/regrets/wants intelligently that he leave.SUBJ

(8) $Roger_i$ ***réalise/dit*** *intelligemment qu'$il_{i/j}$* *part*
R. realizes/says intelligently that he leaves.IND

What (7) vs. (8) suggests is that external arguments have something to say about the semantics of main verbs. If we can identify the common core feature of external arguments that are involved with verbs triggering the subjunctive vs. indicative, we might find which feature distinguishes between the two types of predicates involved in the choice of mood.

I propose that the lexical semantics of these predicates play a crucial role in mood selection. More specifically I show that what regulates the subjunctive is a special feature of the main predicate: the emotive feature. Based on Dowty (1991) and in the spirit of Ramchand (2008a, b), I decompose these predicates into verbal primitives. I show that verbs are composed of a bundle of organized features, and that there are several types of subjunctive-selecting verbs, all involving the emotive feature.[1]

My discussion is guided by the observation that there are predicates with mood alternation, (9), but which do not alternate in terms of veridicality (*comprendre* involves in both cases some kind of veridicality (see Baunaz & Puskás 2014)). Thus, veridicality cannot be the crucial factor for mood choice. However, these verbs vary in meaning: if *comprendre* in (9a) involves some intellectual exercise only, it also involves some kind of empathy in (9b). Therefore, with respect to mood selection, verbs can be classified in three groups in (10):

(9) a. *Rafa comprend que Roger* ***veut*** *redevenir numéro 1*
Rafa understands that Roger wants.IND to be.again number 1

b. *Rafa comprend que Roger* ***veuille*** *redevenir numéro 1*
Rafa understands that Roger want.SUBJ to be.again number 1

(10) a. predicates which only take subjunctive CPs
b. predicates which exclusively take indicative CPs
c. predicates which optionally take one or the other type of CP

Section 2 introduces the framework of nanosyntax and the notion of veridicality developed in Baunaz & Puskás (2014). In Section 3, I argue that verbs selecting the subjunctive are *emotive*, while those selecting the indicative are neutral with

1. Nothing about the role of the subjunctive mood in the semantic composition of the sentence is discussed. See Villalta (2006) for a semantic approach to that issue in line with a Kratzerian approach to modality, for which meaning of an overall sentence has to do with the embedded clause (Kratzer 2013; Moulton 2009, a.o.).

respect to emotivity. I also set grammatical diagnostics in favor of the existence of the two classes. In Section 4, I extend the definition of emotivity to impersonal constructions with modals and adjectives that take subjunctive CPs. Finally I provide a nanosyntactic analysis and propose that the distinction between the two types of predicates lies in their internal syntax: emotive predicates are syntactically and semantically bigger than non-emotive ones.

2. Theoretical background

2.1 Nanosyntax (Starke 2009, 2011; Caha 2009)

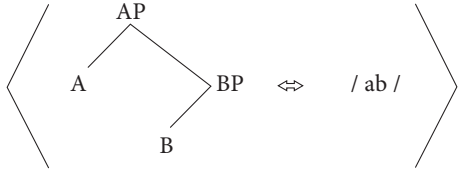
Nanosyntax studies the fine-grained structure of morphemes. It is based on the reasoning that the multiplication of syntactic projections and the subsequent atomization of functional heads impact the architecture and principles of grammar. The consequence is that one single individual morpheme is not considered as realizing one single head, but may span over several of them. As in Cartography, every (morphosyntactic/semantic) feature is a syntactic head; the presence/absence of a property directly affects the size of the structure (Rizzi 1997; Cinque 1999). The features constituting a morpheme are hierarchically ordered according to the universal functional sequence of syntax (= fseq). Because of this, nanosyntax allows for multiple features to be lexicalized at once (= phrasal spellout). There is, thus, no lexicon before syntax; syntax is responsible for building lexical items.

Syntactic structures must be lexicalized by lexical structures. One principle lexicalizing syntactic structure with lexical structure is the *Superset Principle*, as expressed in (11) (from Baunaz & Lander under review, (17)) and illustrated in (12).

(11) The Superset Principle: A lexical tree L can match a syntactic tree S if L is a superset (proper or not) of S.

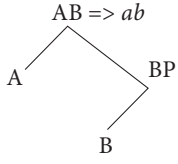
The Superset Principle allows for a single lexically stored structure, L, to 'shrink' in order to match smaller subsets of the structure in the syntax. In (12), L can spell out both syntactic trees S1 [AP [BP]] and S2 [BP]. L matches S1 perfectly, and L is a superset of S2. Both, therefore, spell out as /*ab*/, since this is the phonological form of the lexical entry. Nanosyntax, thus, takes syncretism seriously. Syncretism arises when two or more distinct grammatical functions are spelled out by a single morpheme, i.e. when a lexical entry maps onto a range of syntactic trees, as in (12).

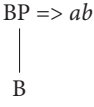
(12) **(L) Lexical entry** (= a lexical tree linked to a phonological form)



Synctactic trees: S1 (=[AP[BP]]) and S2 (=[BP])

(S1) **(S2)**





(credits to E. Lander)

To study the fine-grained structure of languages, nanosyntacticians have at their disposal different tools: (i) compositionality of semantics, (ii) syncretism, and (iii) morphological containment. This paper focusses on (i)–(ii) and shows that they play crucial roles in building up the fseqs of verbs and in determining structural size differences (Sections 3–5). In particular, (i)–(ii) help capturing the apparent ambiguity raised by verbs like *comprendre*, and help us in developing an analysis accounting for the trigger of mood selection.

2.2 Veridicality, non-veridicality, and relative veridicality

Baunaz & Puskás (2014) investigate the notion of 'some individual' in (2) and see how it applies to the French verbs presented in (4) (see also Tóth 2008 for Hungarian). They propose that epistemic factive verbs require that the embedded proposition be true *both* from the point of view of the Speaker and from that of the Subject: they are **strongly veridical**. (13a) shows that *p* must also be true from the Speaker's point of view. In (13a), if Pierre realizes/understands *p*, then *p* must be true in Pierre's knowledge model. In contrast, emotive factive verbs require that the embedded proposition be true from the point of view of the Subject, but not (necessarily) from that of the Speaker. With these verbs, the truth of the embedded proposition is not absolute, but **relative**.[2] If *regrette/comprend* presupposes that x believes that p, it doesn't necessarily follow that p is true (Schlenker 2005; Egré 2008). In fact, *p* is true, but only with respect to the epistemic model of Pierre (and

2. See also Giannakidou (1998).

not (necessarily) with respect to that of the Speaker). Thus, "the shift in the relevant epistemic model (that is, of the Speaker or of the Subject) allows for making different inferences with respect to the truth of the embedded proposition" (Baunaz & Puskás 2014: 246). Note that when *comprendre* selects the subjunctive, it can also trigger a possible shift in the epistemic models of Subject versus Speaker.

(13) P. wrongly believes that Mary is getting married, and…
a. #***realize/comprend*** *qu'elle* *ne reste pas* *célibataire*
realizes/understands that she doesn't stay.IND single
b. ***regrette/comprend*** *qu'elle* *ne reste pas* *célibataire*
regrets/understands that she doesn't stay.SUBJ single

Finally, verbs of saying, of desire, and directive verbs do not embed propositions whose truth must be inferred by the Subject or by the Speaker; they are **non-veridical**.[3] In (14), negating the complements does not yield contradictory statements.

(14) John is convinced that it rains, and…
a. ***veut/suggère*** *qu'il* *pleuve.* (But of course it doesn't rain!)
wants/suggests that it rain.SUBJ
b. ***dit*** *qu'il* *pleut.* (But of course it doesn't rain!)
says that it rains.IND

Hence, the verbs in (10) come in three semantic flavors: strong, relative or non-veridical (Table 1). Also, verbs like *comprendre* behave like either *strong,* or *relative veridical verbs*. Thus what governs the distribution of mood cannot be *veridicality*. Both relative veridical and some non-veridical verbs select the subjunctive (13b), (14a); all veridical and some non-veridical verbs select the indicative, (13a), (14b).

Table 1. French Mood and (non)-veridicality

Verbs	**(non)-veridicality**	**Mood**
Epistemic verb (incl. *comprendre* 'realize')	Strong veridical	Indicative
Emotive factive verbs (incl. *comprendre* 'understand')	Relative veridical	Subjunctive
Verbs of saying	Non-veridical	Indicative
Desire verbs		Subjunctive
Directive verbs		

3. Verbs of saying are veridical for Giannakidou: they select the indicative. (14) shows that semantically they should better be grouped with non-veridical verbs in French.

3. Personal constructions

I propose that the key to mood choice lies in the (lexical) semantic decomposition of the selecting verb (see Puskás 2013). We will see that the distinct semantics of verbs selecting the subjunctive and of those selecting the indicative coincides with systematic grammatical differences, advocating for syntactic (and semantic) decomposition of the verbs in question.

I claim that the matrix external arguments of verbs taking the subjunctive share some thematic feature crucial to mood selection: the *emotive* construal assigned by the emotive feature, which is part of the feature make-up of the verb (cf. lexical entries may map onto a range of syntactic trees in nanosyntax and that syntactic trees are made of hierarchically ordered (submorphemic) features). External arguments of verbs taking the indicative lack this construal because these verbs lack the emotive feature. When the external argument is absent from the structure (as in impersonal constructions with one-argument adjectives), the emotive feature is still part of the feature make-up of these predicates, although it has to be licensed by default by the Speaker (Section 4).

3.1 Decomposing the meaning(s) of verbs

Ramchand (2008a, b) proposes that (dynamic) verbs are decomposed into subparts of the fseq, where "each projection corresponds to a subevent with its own predicational Subject position, and linked by the generalized 'leads-to' or 'cause' relation" (Ramchand 2008b: 118). Each subevent of a predicate is associated to a specific syntactic head. Ramchand proposes three projections, (15): InitP, which introduces the causation event and licenses the external argument (= the initiator, the Subject of 'cause'). The initiator is generally agentive; ProcP, which specifies the nature of the change/process and licenses the entity undergoing change/process (= the undergoer, the Subject of 'process'). ProcP is the dynamic part of the event; ResP, which gives the result state of the event and licenses the entity that comes to hold the result state (= the resultee, the Subject of 'result'). Cause events and resultant events are states, whereas the process event is a process (= VP).

(15) $[_{\text{initP}}$ DP3 init $[_{\text{procP}}$ DP2 proc $[_{\text{resP}}$ DP1 res [XP]]]]

(Ramchand 2008b: 118 (3))

Stative verbs have a smaller structure. Ramchand argues that they consist of a stative event and a rheme (licensed in complement position). Syntactically, stative verbs are vPs only, the specifier of those being the holder of the state, (16):

(16) $[_{\text{vP}}$ $\text{DP}_{\text{holder}}$ v [$\text{DP/NP}_{\text{rheme}}$]] (Ramchand 2008a: 55 (34))

(16) predicts that all stative verbs have a similar structure, i.e. epistemic verbs, verbs of saying, emotive factive verbs and desire verbs should all display the structure in (16), since these three types of verbs are stative verbs. Directive verbs should display the structure given in (15). These verbs contain some causation meaning, i.e. they should involve IntP.

We have hinted in (5)–(8) that these verbs do not behave homogenously when it comes to mood choice. The distinct behaviors of these verbs do not follow the stative vs. non-stative cut. Among the stative verbs, epistemic verbs behave differently from desire verbs. Moreover, non-stative directive verbs behave like (stative) desire and emotive factive verbs. In other words, the structures proposed by Ramchand do not help us to distinguish between verbs taking the subjunctive and verbs taking the indicative.

Another issue raised by an analysis *à la* Ramchand for verbs selecting embedded CP comes from veridicality and selection. Baunaz (2015, 2016) shows that *veridicality* governs the distribution of complementizers in MG, Serbo-Croatian (SC) and French (Fr). The verbs in (4) create different (veridical) domains (see Table 1), and, as such, do not select the same type of complementizer. Crucially similar morphological occurrences of complementizers are not considered as being instances of homophony, but of syncretism. In nanosyntactic terms one lexical entry, with one phonological form, can spell-out three distinct syntactic trees. Therefore, what looks like uniform complementizers (*que, da,*[4] *pu*), corresponds in fact to chunks of structures from the same fseq, selected by the different types of verbs of Table 1. Shaded areas indicate syncretism in Table 2.

Table 2. Declarative complementizers

	Non-veridical verbs	**Strong veridical verbs**	**Relative veridical verbs**
SC	da_1	da_2	što
MG	oti	pu_1	pu^2
Fr	que_1	que_2	que_3

Hence, the stative verbs we are interested in do not behave homogenously with respect to veridicality and complementizer selection either.

For these reasons, I discard Ramchand's analysis, although my analysis will also argue for a strong isomorphism between syntax, semantics, and morphology (Section 5).

4. Like MG *na*, SC *da* under desideratives/directives is a mood particle (see Giannakidou 2009; Socanac 2011).

3.2 Dowty (1991)

Dowty (1991) argues that the semantic roles of arguments should not be viewed as discrete categories, but rather as prototypical concepts (Rosch 1973): the Proto-AGENT/the Proto-PATIENT. Crucially these prototypical concepts have no clear defining boundaries. Rather, they are characterized by a list of contributing properties especially dedicated to Subjects and to objects and must be conceived as lexical entailments imposed by a verb on its arguments, that is, the part that it plays in the event it describes (e.g. whether it is volitional, causal, involves movement, etc.), i.e. Dowty decomposes the notions of agent/patient into more discrete components. (17) corresponds to the list of properties for the Proto-Agent (where *sentience* is a cover term used for a cognitive state, emotion or perception, i.e. meaning *S is conscious of p*):

(17) **Proto-AGENT entailments** (Dowty 1991: 572 (27))
volitional involvement in the event or state
sentience (and/or perception)
causing an event or change of state in another participant
movement (relative to the position of another participant)
(exists independently of the event named by the verb)

The matrix Subjects of verbs taking the indicative share thematic features (Table 3, first-second columns). The matrix external arguments of verbs of saying and epistemic verbs must be sentient, but they are neither causative, nor volitional; these verbs never alternate with a subjunctive CP in French, so they don't exhibit obviation and the external arguments of these verbs can be co-referential with the embedded Subject. None of these predicates involve movement. The matrix external arguments of verbs taking the subjunctive share similar thematic features, too (Table 3, third-fifth columns). The matrix arguments of directive, desire and emotive factive verbs must be sentient. To the exception of emotive factive verbs, they are also volitional; only directive verbs involve causation, and their external argument is a causer.

Table 3. Semantic decomposition of verbs selecting tensed CPs

	Saying	**Epistemic**	**Emotive**	**Desire**	**Directive**
Mood	**Ind.**	**Ind.**	Subj.	Subj.	Subj.
Sentient	**yes**	**yes**	**yes**	yes	yes
Volitional	**no**	**no**	**no**	yes	yes
Cause	**no**	**no**	**no**	no	yes

At this point, nothing, other than the mood of their embedded clause, distinguishes emotive factive verbs from verbs of saying/epistemic verbs.

3.3 Emotive vs. non-emotive sentience

There is, however, evidence for distinguishing between verbs taking the subjunctive and verbs taking the indicative when it comes to their lexical semantics. Following Dowty (1991), I have stated that *sentience* is a cover term used for a cognitive state, emotion or perception, i.e. meaning *S is conscious of p*. I claim that *sentience* can be split into two properties: *emotive sentience* vs. (non-*emotive*) *sentience.* In this section, I give a general overview of what these concepts are, and in Section 4, I elaborate on the notion of *emotive sentience* a bit more.

Emotive sentient verbs include an emotive dimension. They describe the emotive (subjective) reaction of a Subject toward an event, i.e. they range over individuals capable of emotivity. They can be relative veridical or non-veridical (typically, desire verbs, emotive factive verbs, directive verbs). The Subject of these verbs has an emotive (subjective) reaction toward its complement (that is, the object of emotivity) (e.g. *volition, desire, empathy, worry, fear, urge, sadness…*), as paraphrased in (18):

(18) a. Roger suggère que… (directive)
= Roger strongly wants X to do P
b. Roger regrette que… (emotive factive)
= Roger feels sorry that P
c. Roger souhaite que … (desire)
= Roger wants X to

"They express the **subjective value of a proposition** rather than knowledge about it or its truth value" (Kiparsky & Kiparsky 1971: 363, see also Léger 2006; Baunaz & Puskás 2014).

Non-emotive sentient verbs range over verbs that require a high level of cognitive control and/or detachment from any potential emotive experience. They mainly describe a knowledge, a saying. The matrix Subject of these verbs expresses neutral (objective) states/reactions with respect to emotivity, and must be able of mental activity. They can express intellectual/logical/cognitive exercises in that they express knowledge or access to/loss of knowledge (Baunaz & Puskás 2014). Non-*emotive* sentient verbs are non-veridical verbs of saying, or strong veridical verbs.

The emotive and non-emotive readings express a mental state vis-à-vis the event denoted by the matrix verb. Both types denote part of the internal experience of the external argument. Even though the two categories share similar semantic features (like sentience), they nevertheless differ when it comes to the type of psychological state they denote (emotive/subjective vs. non-emotive/neutral/objective).

These readings have grammatical correlates. In addition to Subject obviation and mood choice, the emotive reading coincides with systematic differences in adverb modifications (see also Baunaz & Puskás 2014 for similar examples): (i) degree adverbs can only modify emotive verbs and (ii) agent-oriented adverbs can only modify the Subject of non-emotive verbs.

i. Directive verbs (*insister*), desire verbs (*désirer*) and emotive factive verbs (*regretter*) can be modified by adverbs that modify their degree of intensity (Villalta 2006 for Spanish), like *tellement* 'tremendously', *assez* 'quite', *un peu* 'a little', (19). Intensifiers are incompatible with epistemic verbs (*réaliser*), and verbs of saying (*dire*), (20):[5]

(19) *Stan* ***insiste/désire/regrette*** *tellement* *que* *Roger* *parte*
Stan insists/desires/regrets tremendously that R. leave.SUBJ

(20) **Stan* ***réalise/dit*** *tellement* *que* *Marie* *part*
Stan realizes/says tremendously that Mary leave.IND

ii. Subject-oriented adverbs express an inherent quality of the Subject referent (Ernst 2002). Therefore, depending on their semantics, these adverbs should then be able to modify one or the other type of Subject. So-called "agent"-oriented adverbs (= subset of Subject-oriented adverbs) like *intelligemment* 'intelligently' can only be "modifiers in clauses which expect the Subject referent to think something through" (Kelepouris 2012), and as such they should be compatible with non-emotive Subjects, which are traditionally understood as being *agentive*, (21). Emotive Subjects, which are non-agentive, should be incompatible with these adverbs, except for directive verbs, which involve performative Subjects (i.e. agentive Subject). Yet, these adverbs are neither compatible with desire and emotive factive verbs, nor with directive verbs, (22):

(21) *Jean* ***réalise/observe*** *intelligemment* *que* *Marie* *part*
Jean realizes/observes intelligently that Marie leaves.IND

(22) * *Jean* ***insiste/désire/regrette*** *intelligemment* *que* *Marie* *parte*
Jean insists/desires/regrets intelligently that Marie leaves.SUBJ

The data presented in (19)–(22) show that emotive predicates can all be modified by adverbs that modify their degree of intensity, but cannot be modified by agent-oriented adverbs. Non-emotive predicates cannot be modified by degree adverbs but can be by agent-oriented adverbs.

---

5. *Dire* 'say' can co-occur with *assez*, but when it does, *assez* has a quantity meaning (vs. degree). See Villalta (2006) for discussion and for an account of the possibility of *know very well* in Spanish and English.

I claim that *emotive* predicates must be semantically distinguished from non-*emotive* predicates, in that the external arguments of these predicates express different states: the former expresses an emotive (subjective) reaction, and the latter expresses neutral states (with respect to emotivity) toward the information conveyed by their complements. In Section 3.4 I show that emotive predicates add something more to the neutral meaning.

3.4 Verbs optionally taking indicative or subjunctive CPs: Emotive vs. non-emotive readings

Verbs that optionally take indicative or subjunctive CPs alternate between an *emotive* and a non-*emotive* reading. Emotive verbs are shown to be compatible with comments on the Subject's emotive state but not with comments on his logical state, whereas non-emotive verbs disallow comments on the Subject's emotive state (Baunaz & Puskás 2014).

In (23), the meaning of *comprendre* is very close to the English verb 'realize'. Rafa adds to his intellectual environment the fact that Roger is turning number 1 again. The external argument of *comprendre* is sentient. *Comprendre* expresses an intellective exercise only: the continuation commenting on Rafa's emotive state is infelicitous, (23b). On that reading, *comprendre* is an epistemic verb and denotes "a cognitive activity which results in a change of a person's epistemic state" (Becker 2010). Commenting on Rafa's intellectual state is fine (23a):

(23) *Rafa* ***comprend*** *que Roger redevient numéro 1*
Rafa understands that Roger turns.IND number 1

a. but he is wrong (he = Rafa) (non-emotive)
b. #but it makes him sad (him = Rafa, it = realize) (emotive)

In addition to being an intellective exercise, *comprendre* in (24) involves an empathic attitude of the Subject toward the complement (Becker 2010) and can be paraphrased as in *Rafa empathizes that Roger is turning number 1 again*. The external argument of *comprendre* is also sentient. A continuation such as *but it makes him sad,* i.e. commenting on his emotive experience, is perfect: on that reading, *comprendre* is an emotive verb.

(24) *Rafa* ***comprend*** *que Roger redevienne numéro 1.*
Rafa understands that Roger turn.SUBJ number 1

a. #but he is wrong (he = Rafa) (non-emotive)
b. but it makes him sad (him = Rafa, it = understand) (emotive)

Verbs like *comprendre* vary in meaning: *comprendre* in (23) only involves an intellectual exercise, whereas in (24) it also involves empathy from the main Subject toward the complement clause. In that sense, (23) is closer to the core meaning of the verb *comprendre* than (24), which, in addition to the intellection reading, also involve an emotive construal. The meaning of these predicates influences the mood of the CP-complement they select.

Again I claim that these readings have grammatical correlates (Section 3.3). In addition to mood choice, the emotive reading coincides with systematic differences in terms of adverb modification, (i)–(ii) and obviation (iii).

i. Degree modification can only modify *emotive* verbs, but cannot modify non-*emotive* verbs. This is illustrated, when *comprendre* takes the indicative in (25a) and when it selects the subjunctive, in (25b):

(25) a. **Stan* ***comprend*** *tellement* *que* *Roger* *part*
Stan realizes tremendously that Roger leave.IND

b. ***Stan*** ***comprend*** *tellement* *que* *Roger* *parte*
Stan understands tremendously that Roger leave.SUBJ

ii. "Agent"-oriented adverbs can only modify non-emotive verbs, which display a high level of cognitive control, but cannot modify emotive verbs, (26).

(26) *Jean* ***comprend*** *intelligemment* *que* *Marie* *part/*parte*
Jean realizes/understands intelligently that Marie leaves.IND /*SUBJ

iii. Embedded Subjects under emotive verbs selecting subjunctive CPs show obligatory obviation (28), while embedded Subjects under non-emotive verbs do not, (27):

(27) $Roger_i$ ***comprend*** $qu'il_{i/j}$ *a* *gagné*
R. understands that he has.IND won

(28) $Roger_i$ ***comprend*** $qu'il_{*i/j}$ *ait* *gagné*
R. understands that he has.SUBJ won

The matrix arguments of *comprendre plus* indicative share similar thematic features with verbs of saying and epistemic verbs. They are sentient, but not volitional, nor causative, or emotive, i.e. they resemble non-emotive strong veridical verbs. The matrix arguments of *comprendre plus* subjunctive share similar thematic features with emotive factive verbs, too. They must be sentient and emotive, but neither volitional, nor causative, i.e. it resembles emotive relative veridical verbs.

We have seen that the emotive *comprendre* contains the non-emotive reading (both are sentient, and emotive construals add something to the non-emotive ones).

I claim that the presence or absence of the emotive feature is crucial for the choice of mood (Table 4).[6]

Table 4. French Mood, veridicality and attitude

	Veridicality	**Mood**	**Attitude**
Emotive factive	Relative	Subjunctive	Emotive
Desire	Non-veridical		
Directives			
Epistemic	Strong	Indicative	Non-emotive
Verbs of saying	Non-veridical		

In this section, we have seen that the emotive vs. non-emotive readings have grammatical correlates: in addition to mood choice, they coincide with systematic differences in terms of adverb modification, and Subject obviation. Thanks to apparent 'homophonous' verbs like *comprendre* (see Section 5 for an account), we have shown that the emotive reading contains the non-emotive one and that both readings involve sentience, i.e. emotive verbs involve both the sentient and emotive readings; non-emotive verbs only involve the sentient reading. Our working hypothesis has been that the emotive semantics has an influence on the matrix external argument: emotive sentience verbs describe the emotive (subjective) reaction of a Subject toward an event, i.e. they range over individuals capable of emotivity. In other words, the Subjects of these verbs experience emotive (subjective) reactions toward their complements (= object of their emotivity). This definition is too restrictive though and needs to be loosened up in view of the existence of impersonal constructions involving modal adjectives, as well as one-argument adjectives: these predicates embed subjunctive complements although they have no emotive external arguments (Section 4).

6. Note that *espérer* 'hope' is ambiguous between an emotive (close to *wish* in meaning) and a non-emotive reading (close to *count on, consider*, cf. Trésor de la Langue Française (http://www.cnrtl.fr/definition/espérer)). When *espérer* is interpreted as *wish*, the subjunctive is found, (i), when *espérer* has a non-emotive reading, it commands the indicative.

(i) En vain vous espérez qu'un dieu vous le *renvoie* (…)
unsuccessfully you hope that a god to.you him sends.SUBJ back
'Unsuccessfully you hope that a god sends him back to you' (Racine, PHÈDRE)

4. Impersonal constructions

Modal verbs in French do not select finite embedded clauses, but non-finite ones:

(29) a. *Roger* ***peut*** *gagner Wimbledon*
R. can win Wimbledon

b. **Roger* ***peut*** *que Stan gagne* *Wimbledon*
R. can that Stan win.IND/SUBJ Wimbledon

The only way to get a finite embedded clause with a matrix modal is to construct an impersonal structure, as in (30):[7]

(30) *Il est* ***possible*** *que Roger perde* *Wimbledon*
It is possible that R. lose.SUBJ Wimbledon

Impersonal constructions can also appear with adjectives involving the emotive reading discussed in the preceding sections.[8] When they do so, they are one-argument predicates:

(31) *Il est* ***regrettable/{souhaitable/désirable}/{suggéré/requis}*** *que*
It is regrettable/{desirable}/{suggested/requested} that
Stan parte
S. leave.SUBJ

Note that both *possible* and *emotive adjectives* can be modified by degree adverbs:

(32) a. *Il est tellement* ***regrettable/souhaitable*** *que Stan parte,*
It is tremendously regrettable/desirable that S. leave.SUBJ
que c'en est triste
that it.CL is sad

b. *Il a été tellement* ***demandé*** *que Stan parte,*
It has been tremendously asked that S. leave.SUBJ
que c'en est agaçant
that it.CL is irritating

7. For Léger (2006) *possible* optionally embed subjunctive or indicative CPs. In the former case, *possible* can be interpreted deontically; in the latter, it must be interpreted epistemically. I have encountered only one Speaker who accepts the indicative under *possible*. More work needs to be done to see how/when/if the epistemic reading shows up under *possible* + subjunctive CP.

8. French directive verbs do not have adjectival counterparts ending in *-able* or *-ible*. Passive verbs lacking external arguments are then legitimate in impersonal constructions.

c. *Il est tellement* ***possible*** *que Stan parte, que c'en*
It is tremendously possible that S. leave.SUBJ that it.CL
est agaçant
is irritating

Personal and impersonal constructions are not equivalent (see Léger 2006). In personal constructions, the relevant verb is a two-place predicate, while in impersonal constructions the adjective is a one-place predicate (with no external argument position available). In (32), there is no overt entity experiencing an emotive reaction toward the event denoted by the CP-complement (= the object of emotivity). These predicates, nevertheless, involve some emotive dimension toward the embedded event. I claim that in the absence of an external argument, the Speaker gets some kind of emotive (subjective) role. More specifically, the Speakers in these following constructions give their personal judgments (evaluations) toward the embedded events (from Léger 2006: 219, (97)–(98)):

(33) a. *Il* ***est triste*** *(pour Jean) qu'il n'ait pas d'enfants. Or,*
It is sad (for J.) that he doesn't have.SUBJ any children. Yet
il n'a pas l'air de partager mon avis. Il dit être très
he doesn't seem to share my opinion. He says to be very
heureux sans enfant
happy without children.

b. #*Jean* ***est triste*** *qu'il n'ait pas d'enfants. Or*
Jean is sad that he doesn't have.SUBJ any children. Yet
il n'a pas l'air de partager mon avis/or il n'est pas triste
he doesn't seem to share my opinion/yet he is not sad.

(33b) is contradictory. (33a) is not because in the impersonal construction, Jean himself is not necessarily sad, but the situation that is described by the adjective is and it is the Speaker that somehow witnesses sadness. The adjective has only one argument. In the personal construction, it is the Subject of predication that experiences sadness. The adjective is a two-place predicate. The external argument experiences sadness, but the sentences *yet he doesn't seem to share my opinion/ yet he is not sad* negate this experience. Hence, the contradiction. Similar judgments are given for the pairs in (34)–(35):

(34) a. *Il est* ***regrettable*** *(pour Jean) qu'il n'ait pas d'enfants.*
It is regrettable (for J.) that he doesn't have.SUBJ any children.
Or, il n'a pas l'air de partager mon avis. Il dit être très
Yet he doesn't seem to share my opinion. He says to be very
heureux sans enfant
happy without children.

b. #*Jean* ***regrette*** *de ne pas avoir d'enfant. Or il n'a pas l'air*
Jean regrets to have no children. Yet he doesn't seem
de partager mon avis/or il ne le regrette pas
to share my opinion/yet doesn't regret it.

(35) a. *Il est* ***suggéré/souhaitable*** *(pour Jean) qu'il ait des enfants.*
It is suggested/desirable (for J.) that he has.SUBJ children.
Or, il n'a pas l'air de partager mon avis. Il dit être très
Yet he doesn't seem to share my opinion. He says to be very
heureux sans enfant
happy without children.

b. #*Jean* ***suggère d'/souhaite*** *avoir des enfants. Or, il n'a pas l'air*
Jean suggests/wishes to have children. Yet he doesn't seem
de partager mon avis. Il dit être très heureux
to share my opinion. He says to be very happy
sans enfant
without children.

Similarly with modal adjectives, the Speaker gives a subjective judgment about the complement. Both (36) and (37) involve a deontic interpretation (possibility and/or permission vs. epistemic):

(36) *Il est possible (pour Marie$_i$) qu'elle$_i$ ait des enfants*
It is possible (for M.) that she have.SUBJ children

(37) *Marie peut avoir des enfants*[9]
Marie can have children.

In (36)–(37), either the circumstances, or someone allow Mary to have kids. Léger (2006: 10) claims that deontic *possible,* which embeds either a subjunctive or a non-finite complement, expresses a subjective judgment of the Speaker. The Speaker formulates that the event denoted by the complement can be realized (possibility)

9. (37) also permits the ability reading.

or is permitted. The author adds that "The deontic interpretation of possibility and permission are related to a subjective judgment towards an event. Using deontic *être possible*, the Speaker makes a personal judgment on the possibility of realization of an event, or on its permission" (Léger 2006: 10, my translation). Hence, in that sense, modals like *possible* are emotive adjectives too. In Section 5, I give a (nano)-syntactic analysis of emotive vs. non-emotive predicates.

The examples in (29)–(37) suggest that the notion of "Speaker" needs to enter the discussion. If personal constructions involve emotive reactions of the Subject toward an event (= internal to the clause), impersonal constructions involve the Speaker's feelings (= external subjective personal evaluation) towards an event when no overt external argument is present (i.e. by default). In other words, the definition of emotivity must contain the notion of subjectivity of either the Subject (of emotivity) or the Speaker.

5. The proposal

I propose that there are at least three different classes of predicates selecting the subjunctive, of different structural sizes, with three semantics, feeding three different syntactic behaviors, suggesting that these three classes involve different features. These structures must be distinguished from those of predicates selecting the indicative. The different embedding predicates each select different types of mood, whose properties are tied to emotivity vs. non-emotivity. For reasons of space, my proposal mostly focuses on personal constructions (Section 5.1). Impersonal constructions are also accounted for, though in fewer details (Section 5.2).

5.1 The structures of personal constructions

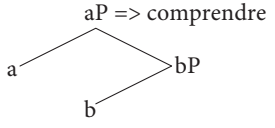
Recall that my discussion is guided by the existence of the alternation emotive vs. non-emotive with verbs like *comprendre*. This alternation coincides with systematic (grammatical) differences suggesting that these predicates involve different features. Since both realizations have the same phonological spell-out, I suggest that they are syncretic items.

In nanosyntax syncretism is a case of a single lexical structure, which can spell out multiple syntactic structures. Recent work has shown that syncretisms reflect structural adjacency, which means that syncretisms inform us about the order of the underlying fseq (Caha 2009).

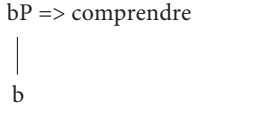
Thanks to the *Compositionality of semantics* tool, I have shown that the non-emotive meaning can be contained within the emotive reading. Non-emotive verbs

are neuter with respect to emotivity. I propose that *comprendre* has only one lexical entry that can spell out two different S-trees, as in (38), where *a* and *b* are syntactico-semantic features (a = emotive, b = sentient). Since the meaning of emotive *comprendre* entails the act of intellection intrinsically displayed by non-emotive *comprendre*, aP is 'eaten up' in (38a), while still available with non-emotive sentient *comprendre*, (38b). aP is then responsible for the emotive reading.

(38) a. (S1) emotive sentient *comprendre*



b. (S2) non-emotive sentient *comprendre*



Thus, in terms of set/super-set terminology, an emotive predicate is semantically and syntactically bigger than a neutral non-emotive predicate.

Extending this reasoning to the verbs in (4), I propose that they all have complex structures and lexicalize a more or less truncated structure. Each property discussed in Dowty 1991 (cause, volitional sentience) is associated with a separate projection, and the same applies to the emotive property as well. Each Spec is a Subject of predication position (as in Ramchand 2008a, b) and licenses the external argument. Verbs of saying and epistemic verbs share with emotive verbs the sentient feature only. Recall that there are three features that are not shared by all the verbs under study. The emotive feature is encoded only in desire, directive and emotive factive verbs; the volitional feature is encoded only in desire and directive verbs, but not in emotive factive verbs and the causal feature is exclusively encoded in directive verbs.

Therefore, *Directive* verbs can be decomposed in various sub-events: causal, volitional, emotive, and sentient, which are each represented by an appropriate projection, (39). The Subject of these verbs also combines these meanings. Semantically, the most prominent reading of these external arguments is that of cause.

(39) $[_{\text{CauseP}}$ DP4 cause $[_{\text{VolitionalP}}$ DP3 vol $[_{\text{EmotiveP}}$ DP2 emo $[_{\text{SentienceP}}$ DP1 sent]]]

Desire verbs are (minimally) decomposed in three stative sub-events, involving volition, emotivity, and sentience, which are each represented by an appropriate projection, (40). The Subject of volition must be sentient and emotive, as well as denotes a volitional reaction toward the object of volition, i.e. the embedded CP.

Volitional Subjects being also emotive and sentient, these meanings compose the external argument of volitional verbs.

(40) $[_{\text{VolitionalP}}$ DP3 vol $[_{\text{EmotiveP}}$ DP2 emo $[_{\text{SentienceP}}$ DP1 sent]]

Emotive factive verbs are (minimally) decomposed in two stative sub-events, involving emotivity and sentience, which are each represented by an appropriate projection, (41). The Subject of emotivity is the emotive Subject, which denotes an emotive reaction toward the object of emotivity, i.e. the embedded CP. Emotive Subjects being also sentient, both meanings compose the external argument of emotive verbs. Semantically, the most prominent readings of emotive arguments is that of emotivity.

(41) $[_{\text{EmotiveP}}$ DP2 emo $[_{\text{SentienceP}}$ DP1 sent]]

Verbs of saying/epistemic verbs is minimally composed of one event, that of sentience. The entity licensing *sentience* is the Subject of sentience, (42):[10]

(42) $[_{\text{SentienceP}}$ DP1 sent]

The features building this fseq are taken to be cumulative, i.e. the structures which are built according to the fseq are in subset-superset relations with one another. Thus, adding an emotive feature to the sentient feature turns the verb into an emotive one (and correlates with the subjunctive in the embedded clause).

5.2 The structures of impersonal constructions

In impersonal constructions involving modals and adjectives, no external argument is realized, yet these adjectives trigger some intrinsic emotive reading (cf. the fact that they can be modified by degree adjectives, (32)). I have shown that these predicates diverge slightly in meaning from those in personal constructions though, in that there is no argument carrying the emotive reaction to the event denoted by the matrix predicate. The emotive (subjective) meaning is attributed by default to the Speaker. In (33)–(37), the Speaker gives a subjective judgment about the complement, i.e. s/he is emotive (see also Léger 2006). The Speaker being emotive, s/he also has to be sentient (since the Speaker is, by definition, sentient).

10. The veridicality feature might be involved in the feature make-up of these verbs. Epistemic verbs are probably bigger than verbs of saying, since they involve a strong veridicality feature. Because this feature plays no role in the choice of mood (Section 2), I haven't included it in the discussion.

I propose that the emotive feature is present, even in the absence of the Subject of predication position. This looks like a genuine case of passivization, where the external (agent) argument is suppressed. In the absence of a referential external argument, the emotive value is assigned by default to the Speaker. Under a strict syntax-semantic mapping, this suggests that the emotive Speaker is somehow represented in the syntax, as a non-overt internal argument (equivalent to *to me* in *it is possible to me that… it is regrettable to me that… it is desirable to me that…*). The details of the analysis go beyond the aim of this paper and will need to be left here for further research.

6. Conclusion

In this paper, I have identified five classes of verbs encoding different semantic and syntactic properties and I have shown that what turns a verb into a subjunctive selecting verb is the emotive feature. The Subject of predicates involving the emotive feature, in addition to the sentient feature, always expresses (at least) an emotive reaction toward the complement CP. In addition to this emotive reaction, it can also express a volitional and a causal state. I have also investigated the semantics of modal, emotive factive, desire, and directive adjectives. These predicates only appear in impersonal constructions and lack external arguments. I have proposed that the emotive (subjective) meaning that is involved with these constructions is assigned by default to a non-overt Speaker, in an internal argument position. In these cases, the Speaker gives a subjective personal judgment (= feeling) about the complement, i.e. the Speaker is emotive.

In the spirit of Ramchand's event decomposition, I have proposed that verbs have complex structures and that the five classes of verbs come with a more or less truncated structure. I conclude that all subjunctive-taking verbs can be analyzed in a common way, despite their semantic diversity.

Acknowledgements

I would like to thank Frédérique Berthelot, Joanna Blochowiak, Ciro Greco, Liliane Haegeman, Christopher Laenzlinger, Eric Lander, Genoveva Puskás, Amélie Rocquet, Tanja Samardžić, Tomislav Socanac, Andrew Weir, all the members of the GIST at UGhent, the audience of LSRL 2014 and two anonymous reviewers for comments, judgments and/or discussions on previous versions of this work. All errors are mine. This research has been partly supported by the Swiss National Foundation (grant: PA00P1_145313 and project: SNF 100015_126861), as well as FWO project 2009-Odysseus-Haegeman-G091409.

References

Baunaz, Lena. 2015. On the various sizes of complementizers. *Probus* 27 (2): 193–236.

Baunaz, Lena. 2016. Deconstructing complementizers in Serbo-Croatian, Modern Greek and Bulgarian. *Proceedings of NELS* 46 (1): 69–77.

Baunaz, Lena, and Genoveva Puskás. 2014. "The Selection of French Mood." In *Selected Papers from the 41th LSRL, Ottawa, May 2011*, ed. Eric Mathieu et Marie-Hélène Côté, 233–253. John Benjamins.

Baunaz, Lena, and Eric Lander. Under review. Nanosyntax: the basics. Ms. Universiteit Gent.

Becker, Martin G. 2010. "Principles of mood change in evaluative contexts: the case of French." In *Modality and mood in Romance*, ed. by Martin G. Becker and Eva-Maria Remberger, 209–234. Berlin: de Gruyter. doi: 10.1515/9783110234343.2.209

Caha, Pavel. 2009. *The nanosyntax of case*. PhD, University of Tromsø

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads*. Oxford: OUP.

Costantini, Francesco. 2009. *Interface Perspectives on Clausal Complementation*. The Case of Subjunctive Obviation. Venezia, Libreria Editrice Cafoscarina.

Dowty, David. 1991. "Thematic Proto-Roles and Argument Selection". *Language* 67: 547–619. doi: 10.1353/lan.1991.0021

Egré, Paul. 2008. Question-Embedding and Factivity. *Grazer Philosophische Studien* 77 (1): 85–125.

Ernst, Thomas. 2002. *The syntax of adjuncts*. Cambridge: CUP.

Giannakidou, Anastasia. 1998. *Polarity Sensitivity as (Non)veridical Dependency*. Amsterdam and Philadelphia: John Benjamins doi: 10.1075/la.23

Giannakidou, Anastasia. 2009. The dependency of the subjunctive revisited: temporal semantics and polarity. *Lingua*, special issue on *Mood,* 883–1908.

Kelepouris, Stavros. 2012. The syntax and semantics of subject-oriented adverbs. *A proposal for a new classification*. Ms. University of Ghent.

Kiparsky, Paul, and Carol Kiparsky. 1971. "Fact." In *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology*, ed. by Danny D. Steinberg and Leon A. Jakobovits, 345–369. CUP.

Kratzer, Angelika. 2013. Modality and the semantics of embedding, slides from presentation at the Amsterdam Colloquium, December 2013.

Léger, Catherine. 2006. *La complémentation de type phrastique des adjectifs en français*. PhD, Université du Québec à Montréal.

Moulton, Keir. 2009. *Natural Selection and the Syntax of Clausal Complementation*, Doctoral dissertation, University of Massachusetts Amherst.

Puskás, Genoveva. 2013. Subjunctives – a Family Business. Talk given at the Séminaire de Recherche, University of Geneva, 16.09.2013.

Quer, Josep. 2009. Twists of mood: The distribution and interpretation of indicative and subjunctive. *Lingua* 119 (12): 1779–1787. doi: 10.1016/j.lingua.2008.12.003

Ramchand, Gillian. 2008a. *Verb Meaning and the Lexicon*. CUP. doi: 10.1017/CBO9780511486319

Ramchand, Gillian. 2008b. Lexical Items in Complex Predications: Selection as Underassociation. *Nordlyd, Tromsø Working Papers in Linguistics.*

Rizzi, Luigi. 1997. "The fine structure of the left periphery." In *Elements of grammar: A Handbook of Generative Syntax*, ed. by Liliane Haegeman, 281–337. Dordrecht: Kluwer. doi: 10.1007/978-94-011-5420-8_7

Rosch, Eleanor H. 1973. Natural categories. *Cognitive Psychology* 4: 328–350. doi:10.1016/0010-0285(73)90017-0

Ruwet, Nicolas. 1984. *Je veux partir / *Je veux que je parte*: on the Distribution of Finite Complements and Infinitival Complements in French. *Cahiers de Grammaire* 7: 75–138.

Schlenker, Philippe. 2005. The Lazy Frenchman's Approach to the Subjunctive. In *Romance Languages and Linguistic Theory 2003*, ed. by Twan Geerts, Ivo Van Ginneken, and Haike Jacobs, 269–309. John Benjamins. doi:10.1075/cilt.270.15sch

Socanac, Tomislav. 2011. Subjunctive in Serbian/Croatian. *GG@G* 7: 49–70.

Starke, Michal. 2009. "Nanosyntax. A short primer to a new approach to language." In *Nordlyd 36.1, special issue on Nanosyntax*, ed. by Peter Svenonius, Gillian Ramchand, Michal Starke, and Knut Tarald Taraldsen, 1–6. CASTL, Tromsø.

Starke, Michal. 2011. Towards an elegant solution to language variation: Variation reduces to the seize of lexically stored trees. Ms. Barcelona, Spain.

Tóth, Enikő. 2008. *Mood Choice in Complement Clauses*. Frankfurt: Peter Lang.

Villalta, Elisabeth. 2006. Context dependence in the interpretation of questions and subjunctives. PhD dissertation, Universität Tübingen.

CHAPTER 2

Towards a unified treatment of Spanish copulas

María J. Arche, Antonio Fábregas and Rafael Marín
University of Greenwich / University of Tromsø / University of Lille 3

This paper sets the basis for a uniform account of the alternation between the two Spanish copulas (*ser* and *estar*) in adjectival and passive clauses. While the copular contrast has been attributed to the different properties of adjectives (e.g. individual vs. stage level) and to an eventive vs. resultative stative dichotomy in passives, this work shows that they all behave alike regarding their temporal interpretation. We derive such uniformity from the syntactic properties of the copulas themselves: *estar* includes an additional component *ser* lacks that makes everything it merges with stative, with particular temporoaspectual properties.

Keywords: copulas, Spanish, *ser*, *estar*, aspect, temporal interpretation, syntax-semantics

1. Introduction

This paper investigates the contrast existing between the two Spanish copulas, *ser* and *estar*. As in other languages, Spanish copulas appear in combination with noun, adjectival and prepositional phrases forming the so-called copular clauses. They also appear as auxiliaries in the structure of progressive and passive clauses. Some of these combinations are privative of only one of the copulas. For example, noun phrases only combine with *ser* (1), and certain prepositional phrases (2) and adjectives (3) disallow for one of the copulas, as shown below. The same goes for the auxiliary in the progressive structure, which is viable only with *estar* (4). However, there are other cases in which both copulas can appear, giving rise to a contrast with semantic differences whose precise description and formalization are still under study. This is the case of a substantial body of adjectives and passive clauses. Examples of each alternation are in (5) and (6).

(1) *Pedro es/*está lingüista*
'Pedro is a linguist.'

DOI 10.1075/rllt.11.02arc

(2) *Pedro *es/está en el jardín*
'Pedro is in the garden.'

(3) a. *Pedro es/*está vegetariano*
'Pedro is a vegetarian.'
b. *Pedro *es/está descalzo*
'Pedro is barefoot.'

(4) *Pedro *es/está haciendo los deberes*
'Peter is doing his homework.'

(5) *Pedro es/está guapo*
'Pedro is handsome.'

(6) *Pedro es/está perseguido*
'Pedro is chased.'

Since it is in these latter cases where minimal pairs can be obtained, we are going to concentrate on them with the aim of contributing to the debate about the features that may be responsible for yielding such a semantic contrast. In particular, we want to set the basis for an account where the copular contrast can be explained in a uniform manner in both adjectival and passive clauses. The distribution of the Spanish copulas with adjectives has been accounted for as unrelated to that with participles, with few exceptions trying to establish a connection (Zagona 2010). The copulas in adjectival clauses are analyzed on different grounds from their distribution as auxiliaries in passives. In the adjectival domain, the most influential explanation has been the one based on the distinction made by Carlson (1977) between individual-level (IL) predicates and stage-level (SL) predicates. According to this view, advocated by many authors (Schmitt 1992; Fernández Leborans 1995, 1999; Fernald 1999; Arche 2006; Marín 2010, a.o.), when *ser* heads the clause, the predicate is interpreted as a property of the individual herself (IL); when *estar* does, the predicate is interpreted as a property of a spatiotemporal variable (SL), rendering the interpretation of the predicate linked to a circumstance. For instance, in (5) above, Juan is understood as a handsome guy with *ser* and as looking handsome for some reason (such as the suit he is wearing or his tan) with *estar*. In contrast, with participles, see (6) above, a different dimension has been used: passives with *ser* have been argued to express eventive passives, while passives with *estar* express stative passives expressing a result (cf. Alcina & Blecua 1975: 904; Marcos Marín 1980: 283, for traditional studies; cf. Mendikoetxea 1999; Kratzer 2000; Keenan & Dryer 2007; Gehrke & Grillo 2009; Zagona 2010).

Since the copulas are the same throughout the adjectival and participial examples, the divergence in the account looks unjustified. Our goal in this article is to motivate the plausibility of an account where the distributions of the copulas are

unified and traced back to the same set of properties, as a way to set the basis for further studies that capture the properties of the two Spanish copulas irrespectively of the kind of complement each one takes. In particular, we will take facts coming from temporal interpretation as an analytical tool to identify common properties in the two sets of contrasts, copular and passive clauses. As we will show, we will attribute the source of different temporal interpretation to the components of the copulas themselves. Since they are the common elements in both copular and passive structures, we understand that all contrasts yielded should be naturally derived from the same root.

In the following section, we will describe some relevant aspects of the temporal interpretation of copular clauses; in particular, the habitual reading that events, but not states, can have in the present tense. In Section 3, we will discuss the temporal properties that *ser* and *estar* passives have by examining two components: the viewpoint aspect properties of each type of passive and the result interpretation that *estar* passives, but not *ser* passives, bear. In Section 4 we will concentrate on *ser* and *estar* + adjectives constructions, and we will compare them with *ser* and *estar* passives. Finally, Section 5 includes some conclusions and loose ends for future research.

2. The temporal interpretation of copular clauses

Our empirical point of departure is the contrasts of sentences in (7) and (8). We will focus on their temporal interpretation to examine what it can reveal regarding the nature of the predicates; in particular we will be guided by Kenny's (1963) observation that event predicates yield a habitual interpretation in the present tense, while states cannot.

(7) a. *Ese escaparate es decorado/admirado*
that shop-window is^{ser} decorated/admired

b. *Ese escaparate está decorado/*admirado*
that shop-window is^{estar} decorated/admired

(8) a. *Juan es guapo/amable*
Juan is^{ser} handsome/nice

b. *Juan está guapo/amable*
Juan is^{estar} handsome/nice

The sentence of (7a), with a combination of *ser* + participle in the present tense, must be interpreted as habitual if the predicate is an event (*decorate*), as the paraphrase in (9a) below suggests. Crucially, the *ser* passive in the present tense cannot

refer to an ongoing process of decorating that is being developed at the moment of utterance, as (9b) shows. To convey such an interpretation, an explicit progressive is needed (9c).

(9) a. *Este escaparate es decorado regularmente*
this shop-window is[ser] decorated regularly

b. ??*El escaparate es decorado en este preciso instante*
the shop-window is[ser] decorated in this precise moment
Intended: 'The shop window is being decorated right now.'

c. *El escaparate está siendo decorado en este preciso instante*
the shop-window is being[ser] decorated in this precise moment

However, if the predicate is a state (*admire*), its typical temporal interpretation is not habitual but the one known as continuous (Arche 2006, 2014; Bertinetto 1994; Boogaart 1999, a.o.). This aspectual interpretation in the present tense gives the interpretation of the situation as true of an interval that includes the speech time and does not involve a set of repeated occasions in which admiration happens, hence the anomaly in (10). Note that the critical reading here is the one where *admire* means 'to hold a high opinion' rather than 'to contemplate'. Under the relevant interpretation, the sentence can depict the fact that everybody holds a high opinion of the shop window, which is understood to hold an interval that includes the speech time.

(10) *Este escaparate es admirado (#regularmente)*
this shop-window is[ser] admired regularly

Thus, it seems that the interpretation of *ser* passives can be either eventive or stative (De Miguel 1999; Jiménez & Marín 2000). Interestingly, this twofold possibility is not available in *estar* passives. None of the *estar* sentences in (7b) have a habitual interpretation; the passive corresponding to the stative predicate (*admire*) is excluded altogether and the one corresponding to the eventive predicate (*decorate*) is interpreted as referring to the moment of utterance, as seen in (11). That is, the *estar* passive always has stative properties:

(11) *El escaparate está decorado en este momento*
the show-window is[estar] decorated in this moment

The same pattern emerges when the predicate combining with the copula is an adjective, as in (8). Adjectival *ser*-clauses exhibit a twofold behavior depending on the adjective. Adjectives such as *handsome* behave as states, as seen in (12), which shows that the interpretation of the sentence is not habitual; that is, it has the properties expected from a stative predicate.

(12) #*Juan es guapo regularmente*
Juan is[ser] handsome regularly

Ser-clauses can also exhibit properties proper of events. As discussed in Lakoff (1970), Stowell (1991), Chierchia (1995), Fernald (1999), Landau (2006), Arche (2006), Oshima (2009) or Fábregas et al. (2013), with evaluative adjectives *ser*-clauses have the habitual interpretation as a possibility, which is a behavior typical of events. See (13).

(13) *Juan es amable normalmente*
Juan is[ser] nice normally

Reference to the utterance time must be accordingly made by using the progressive, as the contrast below suggests:

(14) a. **Juan es amable en este momento*
Juan is[ser] nice in this moment
Intended: 'Right now, Juan is behaving nicely.'

b. *Juan está siendo amable en este momento*
Juan is being[ser] nice in this moment

Interestingly, however, *estar* again resists a twofold interpretation in all cases, and it allows for reference to the utterance time with both *handsome* and *nice*, as shown by the compatibility of the modifier *en este momento* 'at this moment' in (15) and (16):

(15) *Juan está guapo en este momento*
Juan is[estar] handsome in this moment

(16) *Juan está amable en este momento*
Juan is[estar] nice in this moment
'Right now, Juan is behaving nicely.'

Thus, the generalization that emerges from all the examples considered is the following:

(17) *Ser* yields habitual interpretations with both participles and adjectives, depending on their type, while *estar* only yields the one where reference to the utterance time is possible in all cases.

Some questions that become relevant at this point are the following:

a. Why do *ser*-copular clauses allow for a habitual reading, while *estar* ones do not?
b. What is the distinctive contribution of the copula *estar* to the temporoaspectual domain of the clause?

Our proposal is that the temporoaspectual behavior observed in all kinds of copular clauses is due to the properties of the copulas themselves. We propose that the consistent stative patterning of *estar* clauses is due to the syntactic structure *estar* has, more complex than *ser*'s, consisting of a copular element plus an aspectual component that stativizes the final product. Therefore, the working hypothesis we will be entertaining in this paper is that the difference so often discussed between the two Spanish passives can be traced back to the structure of the copulas themselves, just as we will propose is the case for adjectival contrasts, the latter in line with most traditions.

Note that the account we are proposing to explain the contrast between *ser*/*estar* passives is not one of eventive/adjectival passives, as claimed by other authors (see recent discussions in Zagona 2010). We take the appropriateness of agentive and manner adverbial complements shown in the examples below as a proof of the presence of eventive structure. Crucially, such an eventive structure is stative in the case of *estar*.[1]

(18) a. *El escaparate está* *decorado* *por* *un* *empleado*
the shop-window is[estar] decorated by one employee

b. *El escaparate está minuciosamente decorado*
the shop-window is[estar] carefully decorated

We defend that both types of passives have the relevant structure in charge of eventive-related behavior, and that such an eventive structure is stativized later on in the derivation in the case of *estar*, when *estar* is merged, due to its aspect structure.[2]

1. Note that there is no contradiction here, as states are types of events in the sense of eventualities. Events are opposed to mere adjectives here.

2. We steer away from the discussion about the difference eventive (verbal) and plainly stative (and non verbal) cases corresponding to contrasts such as the one between *vaciado* 'emptied' vs. *vacío* 'empty':

(i) *El* escaparate está (*cuidadosamente) vacío (*por un empleado).
the shop-window is[estar] (carefully) empty (by one employee)

(ii) El escaparate está (cuidadosamente) vaciado (por un empleado).
the shop-window is[estar] (carefully) emptied (by one employee)

Although only the latter can be said to be eventive, both cases combine with *estar*. This suggests, again, that eventivity and *estar* as passive auxiliary are not at odds. In this paper we are interested in the contrast between *ser* and *estar* when both are eventive passives: *El escaparate es/está vaciado* 'the show window is[ser/estar] emptied'.

3. Aspect and tense in passive clauses

The generalization originally enunciated by Kenny in terms of the temporal interpretation of predicates in the present tense can be recast in viewpoint aspect terms, since habituality vs. non-habituality, so-called continuity, are two different ways of imperfectivity (Bertinetto & Lenci 2012; Boogaart 1999; Deo 2012; Arche 2006, 2014, a.o.).[3] Why events can obtain a habitual interpretation in the absence of explicit quantifiers, while states cannot, is a question that exceeds the purpose of this paper, and it will suffice to us to record the contrast in interpretation and give a proposal regarding the origin of the pervasive stativity observed with *estar* and, more in particular, its resultative stativity. In what follows, we will discuss the temporal properties that *ser* and *estar* passives have by examining two components: the viewpoint aspect properties of each type of passive and the result interpretation that *estar* passives but not *ser* passives bear.

3.1 The imperfective interpretation

In this section we introduce a finer-grained description of the imperfective interpretations available in each copular clause (including passives) to use it as an analytical tool to probe into their properties. For the theoretical analysis of viewpoint aspect, we follow interval-ordering approaches as proposed by Klein (1994), Demirdache & Uribe-Etxebarria (2000, 2004, 2007) and, more specifically for Spanish, the account proposed by Arche (2006, 2014). According to these approaches, the imperfective is the viewpoint of semantics where a topic time (TT) is understood as included *within* the interval of the whole eventuality. To illustrate, the TT is the interval corresponding to Tim arriving in (19a) and (19c) and to Tim's being young in (19b). All these intervals are interpreted as included *within* the one of John decorating the shop window and being worried.

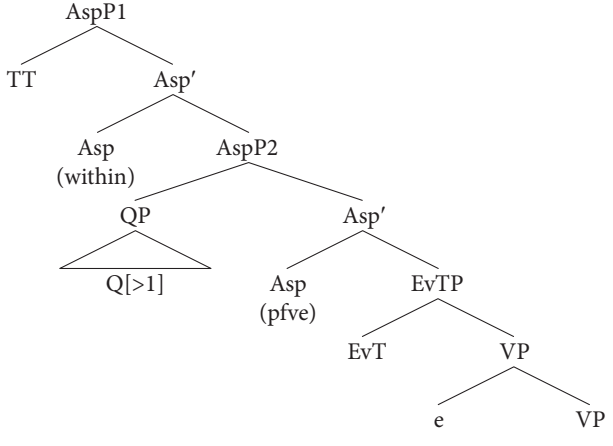
(19) a. When Tim arrived, John was decorating the shop window.
b. When Tim was young, he decorated the shop window.
c. When Tim arrived, John was worried.

However, as acknowledged in the literature (e.g. Boogaart 1999; Verkuyl 1999; Lenci & Bertinetto 2000; Deo 2012, a.o.) not all imperfectives are alike and further distinctions should be made between progressives (which refer to the development of one instantiation of an event), habituals (a plural set of instantiations of the

3. The contrast is equally available in the present and in the past (e.g. as explicit in the morphological dichotomy available in languages such as Spanish between Imperfect and Perfective).

event), and the so-called continuous (plain existential reference to the situation). According to Arche (2006, 2014), all imperfectives share the property of involving the interval-ordering predicate with the value of *within* but differ in other components, which, she argues, are of quantificational nature. Regarding the habitual and the continuous readings of interest for us here, the habitual imperfective contains a quantifier over occasions akin to a distributive quantifier (>1), while the defining characteristic of the continuous would be to lack any cardinality component.[4] The most salient characteristic of the structure of habituality below is that it is conceived as a complex viewpoint: in particular, there is an imperfective predicate (*within*) (observed in the morphology), a quantifier that multiplies the instantiations of the event at hand[5] and a perfective head (i.e. another interval ordering predicate), which captures the fact that the quantifier has a set of finished instantiations of the event as its range (an observation independently found in Lenci & Bertinetto 2000). That is, habituality is felicitous if there has already been a set of finished instantiations whereby the event has taken place.

(20) Habitual structure

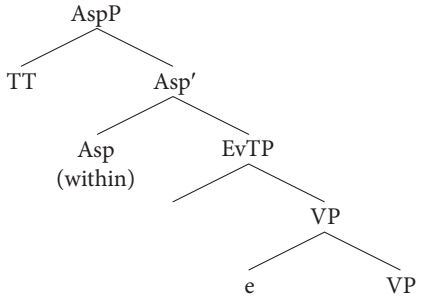


In contrast, the continuous structure in (21) has the same interval ordering predicate (*within*) but does not have a quantifier yielding a multiple number of occasions for the event.

4. The progressive contains a quantifier with the cardinality of one (|1|) as proposed by Verkuyl (1999).

5. The quantifier of the structure corresponds to the quantifiers representing intervals in Stowell's (1993, 1996, 2007) system, so-called *Zeit Phrases*. See Stowell (1993, 1996, 2007) for details and Arche (2006) for discussion on the proportional properties of the habitual quantifier.

(21) Continuous



In the absence of explicit quantifiers, the habitual interpretation is available with eventive predicates while the continuous one has been considered to be typical of states (e.g. Boogaart 1999; Deo 2012). Thus, according to the interpretations described in the section before, we can say that the structure of (20) is in place with those *ser* passives involving an eventive predicate (see (22) below), while the one in (21) with *ser* passives involves a stative predicate (23) and *estar* passives (24).

(22) *El escaparate es decorado (por Juan)* (habitual possible)
the shop-window is[ser] decorated (by Juan)

(23) *El escaparate es admirado* (habitual non possible)
the shop-window is[ser] admired

(24) *El escaparate está decorado por Juan* (habitual non posible)
the shop-window is[estar] decorated by Juan

Ser passives have an additional reading that has not been noticed to the best of our knowledge thus far, which is the one known as *attitudinal*. According to Arche (2014), this reading is the result of the merge of a continuous viewpoint structure to an eventive predicate.[6] This way, a sentence such as (22) can also mean that Juan is in charge of decorating the shop-window, which may be true even in the absence of any actual decoration event. Attitudinal readings are available to event predicates but not to states. This further corroborates the stative character of *estar* clauses, which lack such a reading.

The different viewpoint aspect characteristics of the passives suggest that while *ser* clauses can behave as events or as states, *estar* ones behave as states in a consistent manner. The fact that viewpoint structure is considered to merge to the VP structure predicts that the temporal interpretation derives from both the properties of the VP and the properties of the viewpoint structure. This suggests that the

6. For Arche (2014) there is no selection restriction between event types and viewpoint structure. The interpretation is the result of the merge of the two.

properties of the copula have an impact on the overall temporal interpretation, as is discussed in the next section.

3.2 The aspectual role of the copulas

Our aim in this section is to refine the aspectual characterization of the passive clauses by examining the properties of each one of them.

Regarding *ser* passives, we would like to record the observation that even though the classical idea that they behave as events and seem to preserve the aspectual properties of the event represented in the participle seems overall correct, it is not the case that they are exactly events in the active voice. While it seems the case that the temporal interpretations available with *ser* depend on the type of predicate (continuous if state predicate; habitual if eventive predicate), there is an imperfective reading that is not available for the passive under any circumstance, namely, the progressive. In the active voice in Spanish there is great variation as to whether an event in the present (and marginally in the past) can also be interpreted as referring to the moment of the utterance, that is, as a present progressive. Granted, we can enter a room and ask someone *¿Qué haces?* 'What do you do?' and it will be understood that we are referring to what the person is doing in that precise moment. Something like *Decoro el escaparate* 'I decorate the shop-window' is acceptable. However, in a similar scenario of entering into a room, to a question such as *¿Qué pasa?* 'What happens?' an answer such as *El escaparate es decorado* 'the shop-window is^{ser} decorated' would be markedly deviated, showing that reference to the interval of utterance is not an option for the present tense. In such contexts, of course, the progressive is perfect: *El escaparate está siendo decorado* 'the shop-window is being decorated'.

Estar-passives are always stative clauses, as shown before. More in particular, they are usually described as resultative passives, while *ser* ones are not so. Here we are going to explore the source of such resultative semantics. As a heuristic strategy we will examine the interpretation of temporal modifiers with each of the passives. Consider the sentences below containing a temporal modifier of the type of *a las cinco* 'at five':

(25) a. *El escaparate es decorado a las cinco*
the shop-window is^{ser} decorated at the five

b. *El escaparate está decorado a las cinco*
the shop-window is^{estar} decorated at the five

In the *ser* passive, the modifier is interpreted as referring to the interval in which the decoration happens. In the *estar* passive, the modifier is interpreted as referring

to an interval at which the decoration has already taken place; that is, *at five* designates a moment in which the shop-window is already decorated. These two kinds of interpretations are usually illustrated with the aid of the adverbial *ya* 'already'. Interestingly, the sentences yielded with the addition of the modifier contrast in grammaticality. If *ya* 'already' makes reference to a previous interval where the property did not hold, the degradation of (26) must mean that *ser*-passives cannot make reference to a past interval. Likewise, the fact that this adverbial is allowed in *estar* passives suggests that reference to a prior interval is available.

(26) a. ??*El escaparate ya es decorado*
the show-window already is[ser] decorated

b. *El escaparate ya está decorado*
the show-window already is[estar] decorated

These facts are congruent with the interpretive facts we obtained earlier on in (25) where the modifier could refer to the moment by which the event had taken place only with *estar*.

Given that the only differing element in the minimal pairs is the copula, the null hypothesis is that the source of the ability to refer to a previous interval is *in* the copula. The spirit of this proposal differs from others related to the interpretation of forms containing a participle (e.g. perfect forms), which consider that the different interpretations are rooted in an ambiguity related to the participle. For example, in reference to perfect forms (have + V-ed) Carrasco (1998) argues that participles can be perfective, and refer to the moment in which the event takes place, or perfect, and refer to an interval previous to the event). For others, such as Demirdache and Uribe-Etxebarria (2004), the different interpretations ensue from the different intervals that the adverbial *ya* modifies: the TT, when reference to the moment in which decoration takes place is at hand, and the Event Time, when reference to a moment in which the event was concluded is at hand. Since we cannot extend ourselves here in a discussion about the participles, we will limit ourselves to refer to Stowell (2008), who casts doubt on the widespread assumption that participles encode temporal semantics. In support of this idea, note that the fact that *ya* 'already' is degraded in (26a) means that denotation to an interval of the past cannot be obtained from the participle alone. Furthermore, we want to bring attention to the fact that none of the proposals just mentioned (Carrasco's or Demirdache & Uribe-Etxebarria's) can account for the correlation between each temporal interpretation and the different copula in a natural way.

Thus, the idea we will develop here is the following: the apparent reference to a prior interval is the effect of the resultative semantics present in *estar* passives. This derives from a state aspectual head, which, we propose, is contributed by *estar* itself,

in combination with the process denoted by the main verb. That is, the intuition of a reference to a prior moment is nothing but the presence of a result; results are not a semantic primitive but the outcome of the combination of a predicate that contains a process component and an aspectual head contributing a state, which goes along with Kratzer's (1994) traditional definition. In short, the state of a process is a result. In pursuing this idea, we will analyze two issues: the predicates that can be part of an *estar* passive and the internal composition of *estar*.

3.2.1 *Predicates in combination with estar*

In relation to the first point, it must be noted that not all predicates allow for *estar* passives. Only those that contain a process element produce well-formed sentences in *estar* passives, i.e. activities, accomplishments and achievements, while states are excluded. Table 1 describes the situation and provides examples corresponding to each.[7]

Table 1. Distribution of Spanish copulas according to event types in passives

	States *admirar* 'admire'	**Activities *masticar* 'chew'**	**Accomplishments *decorar* 'decorate'**	**Achievements *encontrar* 'find'**
Ser passive	√	√	√	√
Estar passive	*	√	√	√

(27) States
a. *Juan es admirado*
Juan is[ser] admired
b. **Juan está admirado*
Juan is[estar] admired

(28) Activities
a. *El chicle es masticado*
the gum is[ser] chewed
b. *Este chicle está masticado*
this gum is[estar] chewed

(29) Accomplishments
a. *El escaparate es decorado*
the shop-window is[ser] decorated

7. Note that this generalisation supports an analysis of achievements as essentially accomplishments with a very short process (Ramchand 2008) vs. the account where they lack any process at all (Piñón 1997).

b. *El escaparate está decorado*
the shop-window is[estar] decorated

(30) Achievements
a. *El ladrón es encontrado (cada vez que se escapa)*
the thief is[ser] found (every time he escapes)
b. *El ladrón está encontrado*
the thief is[estar] found

As (28) shows, atelic predicates, with no inherent endpoint in their grammar, can be acceptable in *estar* passives: this fact suggests that the interpretive *result* effects cannot come from the telicity of the predicate that combines with the copula.[8] This also suggests that the result semantics does not derive from the participle semantics, because if it were contained in the participial morphology, the data relating to state verbs (which are excluded from *estar* passives but allowed with *ser* ones) would become unaccounted for. As mentioned earlier on, the proposal we make here is that the result interpretation is obtained syntactically by virtue of the presence of *estar*, which we discuss next.

3.2.2 *The internal composition of estar*

Like most recent traditions, we take it here that the copula *estar* is more complex than *ser*. More specifically, following Hale (1986), Hale and Keyser (2002), and Gallego and Uriagereka (2009), we argue that copulas are heads whose content can be described in the topological terms known as central vs. non-central coincidence. Such topological relations are a universal semantic opposition underlying the predicational, aspectual, modal and complementizer system. For Hale and Keyser (2002), a central coincidence relation defines English copular verb *be*. For Spanish, Gallego and Uriagereka (2009) argue that copula *ser* consists of a copular head and *estar*

8. When no auxiliary of any type is present, the sentences are markedly deviated, which converges with our idea that the endpoint interpretation does not come from the participle, but from additional structure provided by the auxiliaries involved in the different constructions: *have* in the perfect tenses and copular verbs such as Spanish *estar* in the resultative passives.

(i) ??*Conducido el coche, Alonso se quedó tranquilo
driven the car, Alonso SE stayed satisfied

(ii) ??*Arrastrado el perro por la acera, penalizaron al dueño
dragged the dog by the sidewalk, they.punished the owner

(iii) *Decorado el escaparate, Pedro se pudo ir a casa a descansar*
decorated the shop-window, Pedro SE could go to home to rest

(iv) *Temido el zorro, se marchó a otra calle a merodear
feared the fox, SE left to another street to snoop

consists of a copular head plus a component of non-central coincidence nature. Arche (2012) discusses the character of such an additional element and highlights the fact that a finer grained analysis is needed, as non-central coincidence elements can be of two sorts, according to Hale (1986): centripetal, which, in the realm of prepositions, results in those such as *to*, *towards*, and centrifugal, of the type of *from*. Centripetal prepositions are usually involved in dynamic predicates (e.g. *go* to *school*), while centrifugal ones (e.g. *from*) are in resultative constructions. Arche argues that the head expected to be involved within *estar*, if of non-central coincidence nature at all, must have a centrifugal character, since *estar* passives are typically argued to involve resultative semantics. The example below, containing the preposition *de* 'from', illustrates centrifugal aspectual content and the interpretation as a result in the subsequent copular clause, only grammatical with *estar*.

(31) *Vengo de decorar mi habitación. Por fin *es/está decorada*
I.come from decorating my room. At last it.is[*ser/estar] decorated

This perspective leaves us with the following possibilities, none of which are problem free: (i) *ser* simply includes a copular element and *estar* a copular element and an additional head of non-central centrifugal nature. This view presents the copular element somewhat unanalyzed, as opposed to Hale and Keyser's (2002) proposal that copulas incarnate the central coincidence relation themselves, which can be seen as a drawback. (ii) Both copulas are to be analyzed as topological heads, *ser* of central coincidence nature (as Brucart 2010 argues) and *estar* of non-central centrifugal character. This analysis, whereby *estar* intrinsically involves a resultative component, overgeneralizes a resultive interpretation and becomes inappropriate to account for adjectival cases such as *La caja está vacía* 'the box is empty', where the property is not interpreted as a result and no previous process needs to be understood. In what follows we show the empirical benefits of describing *estar* as a predicate of central coincidence, yielding states, coinciding with Hale and Keyser (2002: 218): "central coincidence consistently corresponds to stativity". We argue that stativity crucially yields resultative semantics if a process component is also present in the structure. That is, resultative interpretation is the state of a process, which is in accordance to intuitions and traditional descriptions (e.g. Kratzer 1994). This view has the advantage of allowing us to maintain a semantics for *estar* that can also account for its prototypical combination with PPs headed by prepositions of central coincidence such as *en* 'in', see (32a), and for its presence in progressive forms (32b), which contain a central coincidence preposition in many languages (Talmy 1978; Bybee, Perkins & Pagliuca 1994):

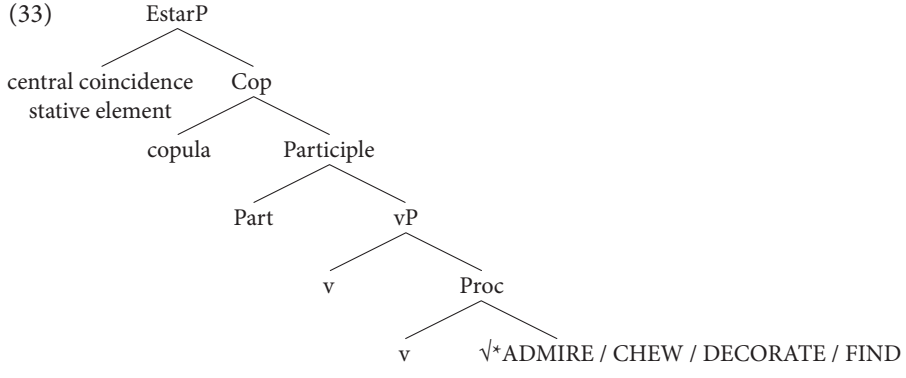
(32) a. *La mesa *es/está en la cocina*
the table is[*ser/estar] in the kitchen

b. *Juan está cazando*
Juan is*ser/estar hunting

Also, for the same reason, the incompatibility between states (*admirar* 'admire') and *estar* (27b) is expected, since a state, with no process component, cannot be *stativized*.

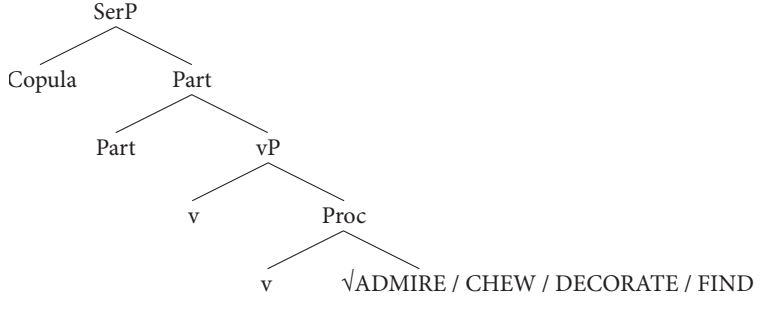
To keep the benefits of the stative character of *estar* we would have to either take Gallego and Uriagereka's (2009) idea of an unanalyzed element (copula) for *ser* or propose a central coincidence character for both copular elements themselves and argue that such an element is reduplicated in the case of *estar* only. We leave the study of this twofold alternative for future work and assume what is schematized below in (33) and (34) for the moment:

(33)



Structure (33) ensures that all *estar* clauses are stative, regardless what the copula combines with. If it combines with a participle, the yielded product will be a result, and, if it combines with an adjective, a stative copular clause. The copula merges with the participle and adds a stative layer onto the structure. Participles from state verbs are arguably excluded due to the impossibility of stativizing a state. In contrast, the configuration for *ser* would be as follows:

(34)



The absence of a stativizing layer makes it possible to merge with all sorts of predicates, including states, which is precluded for *estar*. The structures just proposed would later on in the derivation merge with the higher functional projections of viewpoint aspect and tense. At that point there is no difference between these copular clauses (passives) and any other predicate such as *admirar* 'admire' (a state) or *decorar* 'decorate' (an event); that is to say, the passive clauses are considered by the system as either states (*estar* ones) or events (*ser* ones). If the properties of the passive clauses are due to the properties of the copulas, as we argue, this semantic contribution should extend to other (non passive) copular clauses. In the next section we show that this prediction is borne out.

4. *Ser* and *estar* with adjectives

As introduced at the outset, just as in the case of passives, *ser* yields a twofold possibility regarding temporal interpretation with adjectives, a habitual and a continuous one, while *estar* always gives temporal interpretation in the present where reference to the time of speech is okay, it being aspectually continuous.

(35) a. *Juan es guapo (#normalmente)*
Juan is[ser] handsome (normally)

b. *Juan es amable (normalmente)*
Juan is[ser] nice (normally)

(36) a. *Juan está guapo (#normalmente)*
Juan is[estar] handsome (normally)

b. *Juan está amable (#normalmente)*
Juan is[estar] nice (#normalmente)

Recall that we are not saying that an explicit habitual adverb is excluded with *estar*; the point is that in the absence of explicit adverbials, a habitual interpretation is only available with *ser* followed by adjectives of the evaluative type as discussed in Arche (2006) and Fábregas et al. (2013). A related relevant contrast is the one below. As observed in Arche and Stowell (in progress), a clause such as (37) cannot be used to refer to the interval of the utterance time, while, interestingly, with *estar* the opposite situation holds (38).

(37) #*Juan es cruel con Pedro en este momento*
Juan is[ser] cruel to Pedro in this moment

(38) *Juan está cruel con Pedro en este momento*
Juan is[estar] cruel to Pedro in this moment

This contrast shows that *ser* and *estar* with (evaluative) adjectives pattern exactly the same way as they do with participles in passives. The explanation we give for the properties exhibited in (37) and (38) is the same we have given for the properties of passives. That is, temporal properties are a consequence of the structural properties of the copular verbs. *Estar* includes a head that stativizes whatever it merges with; *ser*, in contrast, lacks such a component, leaving the product with the properties it has. Since evaluative adjectives have the property of either denoting an event or being associated with one somehow (we do not enter into this debate here),[9] the total product exhibits such properties when it is combined with *ser*, and then it is subject to a habitual interpretation. On the other hand, *estar* will take the evaluative adjective and displace its interpretation to one of a state obligatorily, with the same temporoaspectual consequences that become apparent in the case of passives.

5. Some conclusions and loose ends for future research

In this paper we have limited ourselves to show correlations between each of the Spanish copulas and specific temporal interpretations. In particular, a state-like behavior is shown when *estar* is involved and a twofold possibility, state and event-like, when *ser* is. We have derived such behavior from the structure of the copulas themselves and we have argued that this in and by itself can account for the different temporal interpretation of the copular clauses both when they are adjectival and when they form passive clauses.

This is a first step towards a uniform account that can bring together all clauses the copulas are part of. There are, of course, other issues that need to be addressed in future work. We have offered a proposal concerning the source of the event vs. state behavior, but we have not fully undertaken the discussion with respect to the IL/SL characterization of *ser/estar* clauses. In this regard, following seminal insights from Arche (2006), we take it as a working hypothesis that SL-ness patterning of *estar* clauses derives from the presence of an external circumstance, which can represent itself in different ways in the different clauses, as an associated 'event'. In the passives, the external circumstance could be the process that becomes later on stativized in the derivation, while in non-eventive adjectives it could be a perception event or an inferred previous change, among other open possibilities.

9. As discussed in Arche (2006) and Fábregas et al. (2013) evaluative adjectives also have a stative reading in the absence of complements (e.g. *Juan es cruel* 'Juan is^{ser} cruel'), as any other adjective discussed above.

References

Alcina, Juan, and José Manuel Blecua. 1975. *Gramática española*. Barcelona: Ariel.

Arche, María J. 2006. *Individuals in Time. Tense, Aspect and the individual / stage distinction*. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/la.94

Arche, María J. 2012. "A Note on the Individual-Level/Stage-Level dichotomy: how aspectual is it?" *Borealis, An International Journal of Hispanic Linguistics* 1(2): 109–132. doi:10.7557/1.1.2.2385

Arche, María J. 2014. "The construction of viewpoint aspect. The imperfective revisited". *Natural Language and Linguistic Theory* 32(3): 791–831. doi:10.1007/s11049-013-9209-5

Arche, María J., and Timothy A. Stowell. In progress. "Evaluative Dispositional Adjectives and the Individual/Stage Distinction". Ms., University of Greenwich and UCLA.

Bertinetto, Pier Marco. 1994. "Statives, progressives and habituals". *Linguistics* 32: 391–423. doi:10.1515/ling.1994.32.3.391

Bertinetto, Pier Marco and Alessandro Lenci. 2012. "Pluractionality, habituality and gnomic imperfectivity". In *The Oxford Handbook of Tense and Aspect*, ed. by Robert Binnick, 852–880. Oxford, Oxford University Press.

Boogaart, Ronny. 1999. *Aspect and temporal ordering. A contrastive analysis of Dutch and English*. The Hague: Holland Academic Graphics.

Brucart, Josep M. 2010. "La alternancia *ser/estar* y las construcciones atributivas de localización". In *Actas del V Encuentro de Gramática Generativa*, ed. by Alicia Avellana, 115–152. Maestría en Lingüística, Universidad Nacional del Comahue: Comahue..

Bybee, Joan, Revere Perkins and William Pagliuca. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: The University of Chicago Press.

Carlson, Gregory N. 1977. *Reference to Kinds in English*. Ph.D. dissertation, University of Massachussetts, Amherst.

Carrasco, Ángeles. 1998. *La correlación de tiempos en español*. Ph.D. dissertation, Complutense University, Madrid.

Chierchia, Gennaro. 1995. "Individual-Level predicates as inherent generics". In *The Generic Book*, ed. by Gregory N. Carlson and Francis J. Pelletier, 176–224. Chicago & London: University of Chicago Press.

De Miguel, Elena. 1999. "El aspecto léxico". In *Gramática descriptiva de la lengua española*, directed by Ignacio Bosque and Violeta Demonte, 2977–3060. Madrid: Espasa.

Demirdache, Hamida, and Myriam Uribe-Etxebarria. 2000. "The primitives of temporal relations". In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. by Roger Martin, David Michaels, and Juan Uriagereka, 157–186. Cambridge: MIT Press.

Demirdache, Hamida, and Myriam Uribe-Etxebarria. 2004. "The syntax of time adverbs". In *The syntax of Time*, ed. by Jacqueline Guéron and Jacqueline Lecarme, 217–234. Cambridge: MIT Press.

Demirdache, Hamida, and Myriam Uribe-Etxebarria. 2007. "The Syntax of Time Arguments". *Lingua* 117: 330–366. doi:10.1016/j.lingua.2005.07.006

Deo, Ashwini. 2012. "Morphology". In *The Oxford handbook of tense and aspect*, ed. by Robert I. Binnick, 155–183. Oxford: Oxford University Press.

Fábregas, Antonio, Bryan Leferman, and Rafael Marín. 2013. "Evaluative Adjectives are Davidsonian states". *Proceedings of Sinn und Bedeutung* 17: 237–253.

Fernald, Teodore. 1999. "Evidential coercion: Using individual-level predicates in stage-level environments". *Studies in the Linguistic Sciences* 29: 43–63.

Fernández Leborans, María J. 1995. "Las construcciones con el verbo *estar*: aspectos sintácticos y semánticos". *Verba* 22: 253–284.

Fernández Leborans, María J. 1999. "La predicación: Las oraciones copulativas". In *Gramática Descriptiva de la Lengua Española*, ed. by Ignacio Bosque and Violeta Demonte, 2357–2460. Madrid: Espasa Calpe.

Gallego, Ángel, and Juan Uriagereka. 2009. "*Estar* = *Ser* + P". Paper presented at the XIX Colloquium on Generative Grammar, Vitoria, University of the Basque Country.

Gehrke, Berit, and Nino Grillo. 2009. "How to become passive". In *Explorations of Phase Theory: Features, Arguments, and Interpretation at the Interfaces*, ed. by Kleanthes Grohmann, 231–268. Berlin & New York: De Gruyter. doi:10.1515/9783110213966.231

Hale, Kenneth. 1986. "Notes on world view and semantic categories: Some Warlpiri examples". In *Features and Projections*, ed. by Peter Muysken and Henk van Riemsdijk, 233–254. Dordrecht: Foris.

Hale, Kenneth, and Samuel J. Keyser. 2002. *Prolegomenon to a Theory of Argument Structure*. Cambridge (Mass.): MIT Press.

Jiménez, Silvia, and Rafael Marín. 2000. "Por activa y por pasiva". *Actas del IV Congreso de Lingüística General*, Cádiz, 3: 1501–1513.

Keenan, Edward, and Matthew Dryer. 2007. "Passive in the world's languages". In *Clause structure, language typology and syntactic description*, ed. by Timothy Shopen, 325–361. Cambridge: CUP. doi:10.1017/CBO9780511619427.006

Kenny, Anthony. 1963. *Action, emotion and will*. London: Routledge and Kegan.

Klein, Wolfgang. 1994. *Time in language*. London: Routledge.

Kratzer, Angelika. 1994. The event argument and the semantics of voice. Ms., University of Massachusetts, Amherst.

Kratzer, Angelika. 2000. "Building statives". *Berkeley Linguistic Society* 26: 385–399. doi:10.3765/bls.v26i1.1131

Lakoff, George. 1970. *Irregularity in Syntax*. New York: Holt, Rinehart and Winston.

Landau, Idan. 2006. "Ways of being rude". Ms., Ben Gurion University.

Lenci, Alessandro, and Pier Marco Bertinetto. 2000. "Aspect, adverbs, and events: habituality vs. perfectivity". In Speaking of events, ed. by James Higginbotham, Fabio Pianesi, and Achille Varzi (eds.), *Speaking of events*, 245–287. New York: Oxford University Press.

Marcos Marín, Francisco. 1980. *Curso de gramática española*. Madrid: Editorial Cincel.

Marín, Rafael. 2010. "Spanish adjectives within bounds". In *Adjectives: Formal analysis in syntax and semantics*, ed. by Patricia Cabredo and Ora Matushansky, 307–331. Amsterdam: John Benjamins. doi:10.1075/la.153.09mar

Mendikoetxea, Amaya. 1999. "Construcciones con *se*: Medias, pasivas e impersonales". In *Gramática descriptiva de la lengua española*, ed. by Ignacio Bosque and Violeta Demonte, 1631–1722. Madrid: Espasa Calpe.

Oshima, David Y. 2009. "Between being wise and acting wise: A hidden conditional in some constructions with propensity adjectives". *Journal of Linguistics* 45(2): 363–393. doi:10.1017/S0022226709005684

Piñón, Christopher. 1997. "Achievements in an Event Semantics". In *Proceedings of Semantics and Linguistic Theory* 7, ed. by Aaron Lawson and Eun Cho, 273–296. Ithaca (NY): CLC Publications.

Ramchand, Gillian. 2008. *Verb Meaning and the Lexicon: A First Phase Syntax*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486319

Schmitt, Cristina. 1992. "SER and ESTAR: A Matter of Aspect". *NELS* 22: 411–426.
Stowell, Timothy A. 1991. "The alignment of arguments in adjective phrases". In *Perspectives on Phrase Structure: Heads and Licensing [Syntax and Semantics 25*], ed. by Susan Rothstein, 105–135. New York: Academic Press.
Stowell, Timothy A. 1993. "Syntax of tense". Ms., University of California, Los Angeles.
Stowell, Timothy A. 1996. "The phrase structure of tense". In *Phrase structure and the lexicon* [Studies in Natural Language and Linguistic Theory], ed. by Johan Rooryck and Laurie Zaring, 277–291. Dordrecht: Kluwer. doi:10.1007/978-94-015-8617-7_10
Stowell, Timothy A. 2007. "The syntactic expression of tense". *Lingua* 117: 437–463. doi:10.1016/j.lingua.2005.08.003
Stowell, Timothy A. 2008. "Where the Past is in the Perfect". In *Tiempos compuestos y formas verbales complejas* (*Lingüistica Iberoamericana* Vol. 34), ed. by Ángeles Carrasco, 103–118. Madrid: Vervuert.
Talmy, Leonard. 1978. "Figure and ground in complex sentences". In *Universals of Human Language*, edited by Joseph H. Greenberg, 625–649. Stanford: Stanford University Press.
Verkuyl, Henk J. 1999. *Aspectual issues. Studies on time and quantity*. Stanford: CSLI.
Zagona, K. 2010. "Ser and Estar: Phrase structure and aspect". In *Selected Proceedings of Chronos 8 Cahiers Chronos*, ed. by C. Nishida and C. Russi, 303–327. Amsterdam: Rodopi.

CHAPTER 3

How French sheds new light on scalar particles

Isabelle Charnavel
Harvard University

This paper examines the behavior of the French scalar focus-sensitive particles *même*, *quand même*, *ne serait-ce que,* and *seulement* as compared to English *even* and *only*. I first show that French *même* displays a more restricted distribution than *even*; this behavior and that of its antonym *quand même* argue for the scope theory against the ambiguity theory of *even*. Secondly, I demonstrate that the behavior of *ne serait-ce que* and *seulement* reveal the existence of an intrinsic link between *even*-like particles and *only*-like particles. To capture this observation, and more generally the organic relation between scalar particles, I propose a new, parsimonious, theory that builds scalarity, additivity, and exclusivity of scalar particles into a conjunctive or disjunctive meaning.

Keywords: scalarity, focus, *even*, *only*, French

1. Introduction

Theories of the distribution and interpretation of scalar particles such as *even* have not reached any consensus yet as all face problems. Specifically, the ambiguity theories of *even* (i.a. Rooth 1985; Rullmann 1997; Herburger 2000; Schwarz 2005; Giannakidou 2007) and *only* (scalar vs. non-scalar) are uneconomical, the scope theories of *even* (i.a. Horn 1971; Karttunen & Peters 1979; Wilkinson 1996; Lahiri 1998; Guerzoni 2003) have to postulate island violating scope mechanisms, and analyses of 'also only' particles (i.a. Guerzoni 2003) propose presupposition/assertion swapping under negation.

The goal of this paper is to shed new light on how scalar particles partition their meaning domain based on some novel, detailed empirical observations on French. First, it will be shown how the distribution of *même*, which is more restricted than its supposed English counterpart *even*, and that of its colloquial antonym *quand même*, argue for the scope theory (Section 2). Then, we will examine how *seulement*

DOI 10.1075/rllt.11.03cha

is conversely less restricted in its distribution and interpretation than its supposed English counterpart *only*, and how its behavior and that of another particle close to *only*, namely *ne serait-ce que,* demonstrate the existence of an intrinsic link between *only*-like particles and *even*-like particles (Section 3). Based on these novel empirical observations, a new, parsimonious, theory of scalar particles incorporating the scope hypothesis will be proposed, one that derives the *even/only* duality in French, and shows great promise in terms of its crosslinguistic extendability (Section 4).

2. How *même* and *quand même* argue in favor of the scope theory

2.1 Background: The two theories of *even* and their issues

Even is standardly analyzed as a focus sensitive particle that induces ordering of the focus alternatives based on expectedness or likelihood. For instance in (1), *even* implies that the pope was the least likely individual for Julie to invite.

(1) *Julie even invited [the* POPE*]*$_F$

More precisely, *even* takes the whole proposition as argument and associates with the focused constituent *the pope*. As standardly assumed (i.a. Karttunen & Peters 1979), the contribution of *even* is not truth-conditional; it presupposes that the proposition is less likely than its alternatives (scalar presupposition) and some of the alternatives are true (existential presupposition)[1] as formulated in (2) based on Rooth (1985, 1992)'s analysis of focus.

(2) (1) = *even* (p), p: Julie invited the pope
Ordinary meaning of p: $[[\ p]]^{0} = 1$ iff Julie invited the pope
Focal meaning of p: $[[\ p]]^{f} = \{$Julie invited x $|$ x is an individual$\}$
Scalar presupposition: $\forall q\ ((q \in C \wedge q \neq p) \rightarrow p < q)$ where $<$ means "less likely than" and C is a covert variable denoting the set of contextually given alternative propositions, such that $C \subseteq [[\ p]]^{f}$ and $[[\ p]]^{0} \in C$
i.e. the pope is the least likely individual to be invited by Julie
Additive presupposition: $\exists q\ (q \in C \wedge q \neq p \wedge q$ is true)
i.e. Julie invited other people than the pope

1. The quantificational strength (universal vs. existential) of both presuppositions has been debated. I will not take a stand on this issue, which is not crucial for the goal of this paper.

Other orthogonal issues that I will not examine here include whether the scale invoked by *even* should be defined in terms of likelihood, which exact syntactic position *even* occupies, and how it relates to the size of the associate focus constituent.

The debate about *even* is due to its behavior in negative contexts like (3).

(3) *# Julie did not even invite [the* POPE*]*$_F$

As opposed to (1), (3) implies that the pope was the most likely person for Julie to invite; that's why (3) is infelicitous in stereotypical contexts where you invite people you are closest to. Moreover, (3) implies that Julie did not invite anybody. In other words, (3) seems to have reverse presuppositions as compared to (1).

Based on this observation, it has been proposed that *even* takes scope over negation, which derives the correct presuppositions as detailed in (4).

(4) (3) = *even* (p'), p': Julie did not invite the pope
[[p']] ^{0} = 1 iff Julie did not invite the pope
[[p']] ^{f} = {Julie did not invite x | x is an individual}
Scalar presupposition: $\forall q ((q \in C \wedge q \neq p') \rightarrow p' < q)$
i.e. the pope is the least likely individual not to be invited by Julie
Additive presupposition: $\exists q (q \in C \wedge q \neq p' \wedge q$ is true)
i.e. it is also the case that Julie did not invite other people

Specifically, proponents of the scope theory (i.a. Horn 1971; Karttunen & Peters 1979; Wilkinson 1996; Lahiri 1998; Guerzoni 2003; Nakanishi 2006) assume that *even* – just like a Positive Polarity Item (PPI) – takes wide scope over downward entailing (DE) operators.

This theory, however, faces some problems. First, Rullmann (1997) claims that other focus sensitive particles cannot freely scope over negation as exemplified in (5) involving *only*.

(5) *John did not only invite Bill*

(5) cannot mean that it is only the case that John did not invite Bill: *only* cannot outscope the negation. Furthermore, the proponents of the scope theory have to assume that *even* can violate island constraints to raise over DE operators in cases like (6).

(6) a. *Every student who came to class even once will pass the exam*
b. *If John arrived late even once, he will be fired*

In both sentences in (6), the right meaning can be computed only if *even* takes wide scope, thus raising out of the relative clause in (a) and out of the conditional clause in (b), and thereby disobeying island constraints. This is shown in (7).

(7) a. *even (every student who came to class [once]$_F$ will pass the exam)*
Scalar presupposition: 'every student who came to class once will pass the exam' is less expected than 'every student who came to class more than once will pass the exam'.
b. *even (if John arrived late [once]$_P$ he will be fired)*
Scalar presupposition: 'if John arrived late once, he will be fired' is less expected than 'if John arrived late more than once, he will be fired'.

To avoid these problems raised by the scope theory, the proponents of the ambiguity theory (i.a. Rooth 1985; Rullmann 1997; Herburger 2000; Schwarz 2005; Giannakidou 2007) propose two opposite lexical entries for *even* as represented in (8), based on the observation that many languages have a specific form for NPI *even*.

(8) Ambiguous *even* under the ambiguity theory
a. PPI *even*
Scalar presupposition: $\forall q\ ((q \in C \wedge q \neq p) \rightarrow p < q)$
Additive presupposition: $\exists q\ (q \in C \wedge q \neq p \wedge q$ is true)
b. Negative Polarity Item (NPI) *even*
Scalar presupposition: $\forall q\ ((q \in C \wedge q \neq p) \rightarrow p > q)$
Additive presupposition: $\exists q\ (q \in C \wedge q \neq p \wedge q$ is false)

Thus, (3) is not derived by scoping *even* over negation as in the scope theory; on the contrary, *even* is in that case a NPI that needs to be outscoped by negation, and it presupposes that the alternatives are less expected and that some of them are false, i.e. it was less likely for Julie to invite other people than the pope and there are some other people that Julie did not invite. This correctly predicts that (3) is infelicitous in stereotypical contexts.

However, the ambiguity theory also faces problems. First, it is conceptually problematic to suppose that *even* is ambiguous between two opposite meanings depending on the negativity of the context. Secondly, this theory does not capture the fact that the felicity of *even* in the scope of a universal quantifier depends on the content external to the minimal clause in which *even* is base-generated as illustrated in (9).

(9) a. *Every student [who read **even** one paper] will pass the exam*
b. *#Every student [who read **even** one paper] will fail the exam*

For instance, the ambiguity theory would make the same predictions for (9a) and (9b) (i.e. *even* presupposes that it is more likely for a student to read one paper rather than more than one) disregarding the difference of matrix verbs (*pass* vs. *fail*) contrary to facts; in stereotypical contexts, only (9a) is felicitous. The scope

theory, however, correctly captures the difference since it predicts *even* to scope over the whole sentence.

Furthermore, both theories face the same issue in non-DE environments that give rise to the same reading as negative contexts. This reading, which corresponds to the meaning of NPI *even* under the ambiguity theory, will henceforth be called the most-likely reading of *even*, while the reading corresponding to the meaning of PPI *even* under the ambiguity theory will be called the least-likely reading of *even*. The problem arises in modal environments like (10), which are standardly assumed to be upward monotone, and with non-monotone quantifiers as in (11).

(10) a. *Show me even one party that cares for the people*
b. *To pass, John needed to prove he attended the lectures even once*
c. *The band hopes to someday make even one video of that quality*
d. *John is glad that Mary arrived on time even once* (Crnič 2011)

(11) *Exactly four people in the whole world will open this dissertation even once* (Crnič 2011)

Under the scope theory, there is no DE operator over which *even* can take scope; under the ambiguity theory, there is no licenser for NPI *even*, so that both theories incorrectly predict *even* to have the least-likely reading in these cases. In fact, all the propositions in (10) and (11) are presupposed to be more likely than their alternatives; for instance, (10a) presupposes that it is more expected to show one party that cares for the people than more than one such party. Furthermore, there does not seem to be any additive presupposition in those cases: for example, the imperative in (10a) imposes no additional requirement on the addressee other than to show the speaker one political party that cares for the people.

In sum, both the scope theory and the ambiguity theory of *even* are challenged by several problems, some of which are common to both theories. The meaning of *even* according to its distribution is summarized in Table 1: the columns correspond to different contexts of occurrences of *even*, i.e. positive, anti-additive,[2] other DE, and modal/non monotone environments, while the two lines indicate the two readings of *even* – most-likely and least-likely – according to the position of the focus associate on the likelihood scale. While the scope theory captures the most-likely readings of *even* by scoping it over DE operators, thereby violating island constraints in some cases, the ambiguity theory postulates two opposite meanings of *even*; moreover, both theories fail in predicting the reading in the shaded cell.

2. Zwarts (1998) characterizes anti-additive operators (such as the sentence negation or the quantifier *no*) as a subset of DE operators that licence strong NPIs. Anti-additive functions satisfy the following definition: f is anti-additive iff $f(A \wedge B) \leftrightarrow f(A) \wedge f(B)$.

Table 1. Distribution and interpretation of *even*

Position on scale of focus associate/Context of occurrence	Positive	Anti-additive	Other DE	Modal, non monotone
Top e.g. 'invite the pope'	least-likely *even*		least-likely *even*	least-likely *even*
Bottom e.g. 'invite one's best friend'		most-likely *even*	most-likely *even*	most-likely *even*

2.2 French *même*

Même is usually assumed to behave like *even* (i.a. Gast & Auwera 2011; Crnič 2011). But in fact, *même* differs from *even* in crucial ways. In particular, it does not raise the problems that *even* raises under the scope theory, which argues for the scope theory against the ambiguity theory.

First, *même* is not acceptable under the most-likely reading in the environments that are problematic for both theories of *even*, namely in modal and non-monotone contexts. The French equivalents of (10)–(11) are indeed not felicitous in stereotypical contexts as shown in (12)–(13), because they only give rise to the least-likely reading of *même*.[3]

(12) a. ??*Montre-moi* ***même*** *un parti qui se soucie du peuple*
show to-me *même* one party that REFL cares of-the people
'Show me even one party that cares for the people.'

b. ??*Pour passer, Jean devait prouver qu' il avait assisté aux cours* ***même*** *une fois*
for pass John had-to prove that he has attended at-the course *même* one time
'To pass, John needed to prove he attended the lectures even once.'

c. ??*Le groupe espère un jour faire* ***même*** *une vidéo de cette qualité*
the band hopes a day make *même* one video of this quality
'The band hopes to someday make even one video of that quality.'

3. Least-likely readings of *même* are acceptable in these environments as exemplified below.

(i) *Invite même le maire!*
invite *même* the mayor
'Invite even the mayor!'

As opposed to sentence (12), sentence (i) involving an imperative is felicitous, e.g. in a situation where the addressee wants to invite a lot of people to celebrate the opening of his/her new company.

d. ??*Jean est content que Marie arrive à l' heure* ***même*** *une fois*
John is happy that Mary arrives at the time *même* one time
'John is glad that Mary arrived on time even once.'

(13) ??*Exactement quatre personnes dans le monde entier vont ouvrir cette*
exactly four people in the world whole will open this
these ***même*** *une fois*
dissertation *meme* a time
'Exactly four people in the whole world will open this dissertation even once.'

Furthermore, French *même* is not acceptable under the most-likely reading in the counterparts of (6), namely in (14), i.e. in environments that lead to island violation under the scope theory and are predicted to be suitable environments for NPI *even* under the ambiguity theory.

(14) a. ??*Tout étudiant qui est venu en cours* ***même*** *une fois réussira*
all student who is come in course *même* one time will-pass
l' examen
the exam
'Every student who came to class even once will pass the exam.'

b. ??*Si Jean arrive en retard* ***même*** *une fois, il sera viré*
if John arrives in late *même* one time he will-be fired
'If John arrives late even once, he will be fired.'

In sum, *même* exhibits most-likely readings only in anti-additive contexts as shown in (15): the anti-additive operators *ne pas* 'not' and *sans* 'without' license the most-likely reading of *même* as opposed to the non anti-additive operator *peu* 'few'.

(15) a. *Paul* ***n'*** *a même* ***pas*** *invité son meilleur ami*[4]
Paul NEG has *même* not invited his best friend
'Paul didn't even invite his best friend.'

b. *Paul est parti* ***sans*** *même dire au revoir*
Paul is left without *même* say goodbye
'Paul left without even saying goodbye.'

4. The relative surface order of *même* and the negation can be reversed for some French speakers as in (ii):

(ii) %*Paul n' a pas même invité son meilleur ami*
Paul NEG has not *même* invited his best friend
'Paul didn't even invite his best friend.'

c. ??***Peu*** *d' étudiants sont même venus en cours*
few of students are *même* come in course
'Few students even came to class.'

The meaning of *même* depending on the contexts of occurrence is summarized in Table 2.

Table 2. Distribution and interpretation of *même*

Position on scale of focus associate/ Context of occurrence	Positive	Anti-additive	Other DE	Modal, non monotone
Top e.g. 'invite the pope'	*même*		*même*	*même*
Bottom e.g. 'invite one's best friend'		*même*		

Therefore, *même* is well-behaved under the scope theory: no island-violating mechanism needs to be postulated since *même* only takes scope over negation in cases like (15a–b), as represented in (16).

(16) a. ***même*** *(Paul* ***n'*** *a* ***pas*** *invité son meilleur ami)*
même Paul NEG has not invited his best friend.

b. ***même*** *(Paul est parti* ***sans*** *dire au revoir)*
même Paul is left without say goodbye.

Thus, no homonymy has to be assumed for *même* as in the ambiguity theory of *even*: *même* simply corresponds to PPI *even*, and the most-likely reading obtains by scoping *même* over negation.

(17) Contribution of *même* (first version)
Scalar presupposition: $\forall q ((q \in C \wedge q \neq p) \rightarrow p < q)$
Additive presupposition: $\exists q (q \in C \wedge q \neq p \wedge q \text{ is true})$

2.3 French *quand même*

French colloquial *quand même*[5] further argues for the scope theory against the ambiguity theory, which would require more homonymy and synonymy to capture its behavior. Indeed, *quand même* displays the opposite behavior of *même*:

5. French *quand même* also has a concessive use as illustrated in (iii).

(iii) *Caroline est malade, mais elle est quand même venue*
Caroline is sick but she is *quand même* come
'Caroline is sick, but nonetheless she came.'

it exhibits a most-likely reading in positive contexts and a least-likely reading in negative contexts.

(18) #*Paul a quand même invité le pape*
Paul has *quand même* invited the pope
'≈At least, Paul invited the pope.'

(19) *Paul n' a quand même pas invité le pape*
Paul NEG has *quand même* not invited the pope
'≈At least, Paul did not invite the pope.'

(20) *Paul a quand même invité son meilleur ami*
Paul has *quand même* invited his best friend
'≈At least, Paul invited his best friend.'

(21) #*Paul n' a quand même pas invité son meilleur ami*
Paul NEG has *quand même* not invited his best friend
'≈At least, Paul did not invite his best friend.'

In positive contexts like (18) and (20), *quand même* implies that the alternatives are less likely; that is why it can felicitously associate with *son meilleur ami* 'his best friend', but not with *le pape* 'the pope', in stereotypical contexts. By contrast, *quand même* implies that the alternatives are more likely in negative contexts like (19) and (21); that's why conversely, it can felicitously associate with *le pape* 'the pope', but not with *son meilleur ami* 'his best friend'. In other words, *quand même* exhibits the exact opposite behavior of *même* with respect to its presuppositions.

This means that under the ambiguity theory, we would have to postulate the existence of another pair of opposite lexical entries, namely PPI *quand même* and NPI *quand même*, which would be synonymous with NPI *even* and PPI *even*, respectively. This clearly goes against parsimony. On the other hand, the scope theory directly captures the facts: we simply need to assume that *quand même* is the antonym of *même* as formulated in (22), and is a PPI scoping over DE operators, just like *même*.

(22) Contribution of *quand même* (first version)
Scalar presupposition: $\forall q\ ((q \in C \wedge q \neq p) \rightarrow p > q)$
Additive presupposition: $\exists q\ (q \in C \wedge q \neq p \wedge q \text{ is false})$

This correctly predicts the distribution of *quand même* summarized in Table 3, given that *quand même* only exhibits a most-likely reading in DE contexts other

The meaning of *quand même* in (18)–(21) may be historically derived from its concessive meaning; for instance, (20b) could be understood as 'Paul didn't invite the others, but he nevertheless invited his best friend'.

than anti-additive contexts as in (23) and in modal and non-monotone environments as in (24).

(23) *Tout étudiant qui est venu en cours quand même une fois réussira*
all student who is come in course *quand même* one time will-pass
l' examen
the exam
'≈Every student who at least came to class once will pass the exam.'

(24) *Jean est content que Marie arrive à l'heure quand même une fois*
John is happy that Mary arrives at the time *quand même* one time
'≈John is glad that Mary at least arrived on time once.'

Table 3. Distribution and interpretation of *quand même*

Position on scale of focus associate/ Context of occurrence	**Positive**	**Anti-additive**	**Other DE**	**Modal, non monotone**
Top e.g. 'invite the pope'		*quand même*		
Bottom e.g. 'invite one's best friend'	*quand même*		*quand même*	*quand même*

To wrap up this section, the behaviors of French *même* and *quand même* argue for the scope theory against the ambiguity theory as they do not raise the problems posed by *even* under the scope theory (no island violating scope mechanism has to be postulated), and they make the problems of the ambiguity theory worse because more homonymy and synonymy have to be postulated.

3. How French scalar particles reveal an intrinsic link between *even* and *only*

French presents at least two other scalar particles, namely *ne serait-ce que* and *seulement*. They shed further light on scalar particles in revealing an intrinsic link between *even*-like particles and *only*-like particles. These additional new empirical observations will lead us to the building of a new theory of scalar particles in Section 4.

3.1 *Ne serait-ce que*

Ne serait-ce que[6] (literally 'were it only') occurs in contexts where *even* exhibits most-likely readings; in particular, it complements *même* in displaying the most-likely reading where *même* only exhibits the least-likely reading as shown in Table 4.

Table 4. Distribution and interpretation of *ne serait-ce que* as compared to *even* and *même*

Position on scale of focus associate/ Context of occurrence	**Positive**	**Anti-additive**	**Other DE**	**Modal, non monotone**
Top e.g. 'invite the pope'	least-likely *even même*		least-likely *even même*	least-likely *even même*
Bottom e.g. 'invite one's best friend'		most-likely *even même* ***ne serait-ce que***	most-likely *even* ***ne serait-ce que***	most-likely *even* ***ne serait-ce que***

This is first the case in modal and non-monotone environments. We observed in (12)–(13) that the French equivalents of (10)–(11) with *même* are not felicitous because they only give rise to a least-likely reading. Nevertheless, they become as felicitous as (10)–(11) if we replace *même* with *ne serait-ce que* as exemplified in (25) and (26).

(25) *Montre-moi ne serait-ce qu' un parti qui se soucie du peuple*
show to-me *ne serait-ce que* one party who REFL cares of-the people
'Show me even one party that cares for the people.'

(26) *Exactement quatre personnes dans le monde entier vont ouvrir cette*
exactly four people in the world whole will open this
thèse ne serait-ce qu' une fois
dissertation *ne serait-ce que* one time
'Exactly four people in the whole world will open this dissertation even once.'

Moreover, *ne serait-ce que* is the counterpart of *even* in DE contexts that trigger island violation under the scope theory and do not license *même*, e.g. in conditional clauses as exemplified in (27), which is the counterpart of (6b).

6. It is mentioned as *ne fût-ce que* in Gast and Auwera (2011) and Crnič (2011). *Ne fût-ce que* (including the imperfect subjunctive *fût* of *être* 'be' instead of the conditional *serait*) is even more formal than *ne serait-ce que* and is very rarely used.

(27) *Si Jean arrive en retard ne serait-ce qu' une fois, il sera viré*
if John arrives in late *ne serait-ce que* one time he will-be fired
'If John arrived late even once, he will be fired.'

Ne serait-ce que can also be used in anti-additive contexts as in (28) involving *jamais* 'never'.

(28) *Paul n' a* ***jamais*** *eu ne serait-ce qu' un ami*
Paul NEG has never had *ne serait-ce que* one friend
'Paul has never had even one friend.'

In sum, *ne serait-ce que* seems to behave like a NPI inducing the same scalarity presupposition as *quand même*. Moreover, it occurs in contexts that do not trigger any additive presupposition (cf. 'at least') as we discussed in the case of (10a).[7]

(29) *Ne serait-ce que* (first version)
Scalar presupposition: $\forall q\ ((q \in C \wedge q \neq p) \rightarrow p > q)$

Ne serait-ce que is, thus, the reverse of *même* in many respects: it implies opposite scalarity effects and scopes under negation. That's why *ne serait-ce que* turns out to give rise to the same readings as *même* in anti-additive contexts like (28), where *ne serait-ce que* and *même* have both opposite scalarity presuppositions and opposite scopes as represented in (30).

(30) *Paul n'a jamais eu ne serait-ce que/même un ami*
a. ***même*** (Paul has **never** had one friend)[8]
b. **never** (***ne serait-ce que*** (Paul has had one friend))

However, *ne serait-ce que* and *même* exhibit opposite meanings in other DE contexts such as (31); *même* cannot scope over the DE operator in this case because of island constraints, so that *même* and *ne serait-ce que* have the same scope and

7. The absence of additive inference in the case of *ne serait-ce que* clearly appears in positive modal contexts such as the following.

(iv) *Paul voudrait inviter ne serait-ce que Virginie*
Paul would-like invite *ne serait-ce que* Virginia
'Paul would like to invite at least Virginia.'

(iv) does not necessarily imply that Paul wants to invite other people than Virginia. Note that if (28) implies that Paul didn't invite anybody, it is because the alternatives ('Paul has never had more than one friend') are logically implied by the proposition ('Paul has never had one friend'). This is true in the absence of *ne serait-ce que*.

8. For the reasons stated in footnote 4, more speakers accept this sentence with *même* if *même* surfaces before the negation, i.e., before *jamais* 'never' (*Paul n'a même jamais eu d'ami*).

thus opposite scalarity effects due to their meanings. That's why (31b) is felicitous as opposed to (31a).

(31) a. cf. (14) # ***Si*** *Jean arrive en retard* ***même*** *une fois, il sera viré.*
if (***même*** (Jean came late once))
b. cf. (27) ***Si*** *Jean arrive en retard* ***ne serait-ce qu'****une fois, il sera viré.*
if (***ne serait-ce que*** (Jean came late once))

This link between *ne serait-ce que* and *même* interestingly reveals an intrinsic relation between *only*-like particles and *even*-like particles. *Ne serait-ce que,* indeed, includes the exclusive particle *ne … que* close in meaning to 'only'. The behavior of French *seulement,* also close to 'only', makes that link even more striking.

3.2 *Seulement*

French *seulement*[9] is standardly assumed to behave like *only* as suggested by (32).

(32) *Paul a seulement invité un ami*
Paul has *seulement* invited one friend
'Paul only invited one friend.'

However, the following corpus examples reveal an interesting new fact about *seulement*: *seulement* can induce the most-likely reading of *even* in negative contexts. In fact, *seulement* could be replaced with *même* in anti-additive environments like (33)–(34) and with *ne serait-ce que* in other DE environments like (35) without any perceptible change in meaning.

(33) *Un peu plus, tu allais être venu à Paris sans* ***seulement*** *avoir*
a few more you went be come to Paris without *seulement* have
vu notre petit!
seen our child
'You almost came to Paris without even seeing our child!'
(Martin du Gard, les *Thibault*; 1940: 793)

---

9. The adjectival counterpart of *seulement*, namely *seul*, behaves similarly:

(v) *Paul a invité un seul ami*
Paul has invited one *seul* friend
'Paul only invited one friend/Paul invited a single friend.'

Other French expressions corresponding to English *only* include *ne…que*, *rien …que* and *juste*, which are mainly used in colloquial French. They cannot be treated here for space reasons given that their behavior is not identical to that of *seulement/seul*.

(34) *Je crois bien que je ne te donnerai plus rien. Pas*
I believe well that I NEG to-you will-give anymore nothing not
***seulement** ça!*
seulement that
'I think I will not give you anything any more. Not even that!'
(Balzac, *Eugénie Grandet*; 1834: 196)

(35) *La vie est trop courte pour qu' on puisse s'embêter pendant*
the life is too short for that one can REFL bother during
***seulement** une heure.*
seulement a hour
'Life is too short to get bored even for one hour.' (Georges Courteline)

Thus, *seulement* seems to exhibit two readings: one similar to *only* in all contexts, and one similar to *even* (under the most-likely reading) in DE environments.[10] This is illustrated in (36)–(37) and represented in Table 5.

(36) *Luc n' est pas seulement venu une fois.*
Luke NEG is not *seulement* come a time
Only-Meaning: Luke didn't only come once.
Even-Meaning: Luke didn't even come once.

(37) *Est-ce que Luc est seulement venu une fois?*
Q Luke is *seulement* come a time
Only-Meaning: Did Luke only come once?
Even-Meaning: Did Luke even come once? (with negative bias)

10. This reading sounds formal to most native speakers of French. This is not the case when the adjectival counterpart *seul* is used instead of *seulement*: in (vi), the *even*-meaning is not restricted to high register French.

(vi) a. (cf. 36) *Luc n' est pas venu une seule fois*
Luke NEG is not come a *seul* time

b. (cf. 37) *Est-ce que Luc est venu une seule fois?*
Q Luke is come a *seul* time

Table 5. Distribution and interpretation of *seulement* as compared to *even*, *même* and *ne serait-ce que*

Position on scale of focus associate/Context of occurrence	Positive	Anti-additive	Other DE	Modal, non monotone
Top e.g. 'invite the pope'	least-likely *even même*		least-likely *even même*	least-likely *even même*
Bottom e.g. 'invite one's best friend'	***seulement-only***	most-likely *even même* *ne serait-ce que* ***seulement-only*** ***seulement-even***	most-likely *even* *ne serait-ce que* ***seulement-only*** ***seulement-even***	most-likely *even* *ne serait-ce que* ***seulement-only***

3.3 Crosslinguistic link between *even* and *only*

The link between *only*-like and *even*-like particles that is revealed by these new empirical observations on French is observed crosslinguistically: in several languages, the particle used to express most-likely readings of *even* in DE contexts contains an *only*-like particle. This is the case of Italian *anche sole/solanto*, German *auch nur*, Slovak *i len*, Czech *i jen* and Dutch *ook maar*, which literally mean 'also only'; Japanese *dake-demo* and Dutch *zelfs maar*, which literally mean 'even only'; and Spanish *tan solo/solamente* and Catalan *tan sols*, which literally mean 'so only' (i.a. Hoeksema & Rullmann 2001; Guerzoni 2003; Schwarz 2005; Nakanishi 2006; Gast & Auwera 2011; Crnič 2011). Also, the Straits Salish particle *ʔal* has been reported to behave like *only* in positive contexts and like *even* in negative contexts (see Shank 2002; Guerzoni 2003).

To account for the association of an *only*-like particle with an *also*-like particle in negative contexts to induce the most-likely reading of *even*, Guerzoni (2003) assumes that *only*-like particles are scalar and in certain conditions, presupposition and assertion can be swapped.

The scalar reading of *only* can be observed in examples such as (38), as has been argued by i.a. Lerner and Zimmerman (1981), König (1991), Klinedinst (2004), Beaver and Clark (2008).

(38) *Bill only has a BA*

(38) implies that having a BA is lower than the alternatives (e.g. having a MA or a PhD) on a significance scale. Such data motivated the hypothesis of a specific lexical entry for scalar *only* as formulated in (39) based on Guerzoni (2003: 173). Note that

whether the scalar presupposition is always part of the meaning of *only* is a matter of controversy (see discussion in e.g. König 1991).

(39) Scalar *only* defined for *only* (p) such that p is a proposition
  a. Factivity Presupposition: p(w) = 1
  b. Scalarity Presupposition: $\forall q \in C\ [q \neq p \rightarrow q <_{\text{likely/insignificant}} p]$
     i.e. all focus alternatives to p are more significant
  c. Exclusivity Assertion: $\forall q \in C\ [q \neq p \rightarrow q(w) = 0]$
     i.e. the focus alternatives to p are false

Based on this lexical entry, Guerzoni (2003) analyzes the meaning of *also.only*-particles (e.g. German *auch nur*) as the combination of the meaning of *also* and the meaning of *only*. This compositional analysis explains why *also only* is unacceptable in affirmative contexts: the exclusivity conveyed by *only* and the additivity presupposed by *also* (see (40)) are incompatible.

(40) *Also* (p)
  Additivity Presupposition: $\exists q\ [q \in C \wedge q \neq p]$

In negative contexts, the clash can be resolved under the following additional assumptions: *also* in *also only,* but not *only*, can outscope the DE expression, and the factivity presupposition and exclusivity assertion of *only* are swapped; *only*[11] is thus unspecified between the lexical entry in (39) and that in (41).

(41) $Only_2$ (p)
  a. Exclusivity Presupposition: $\forall q \in C\ [q \neq p \rightarrow q(w) = 0]$
  b. Scalarity Presupposition: $\forall q \in C\ [q \neq p \rightarrow q <_{\text{likely/insignificant}} p]$
  c. Factivity Assertion: p(w) = 1

This solution is illustrated with German *auch nur* in (42) (see Guerzoni 2003 for the details of the computation).

(42) *Niemand hat auch nur die Marie getroffen.*
  nobody has also only the Marie met
  'Nobody even met Mary.'
  Scope: ***auch*** (***niemand*** ~~auch~~ ***nur*** hat die Maria getroffen)

In sum, under this view we need to postulate an ambiguity for *only* (between (39) and (41), and potentially also between non-scalar *only* and scalar *only*) to account for the *even*-meaning of *only*-like particles in negative contexts.

11. *Only* is here meant as a generic *only*-particle. Guerzoni (2003) does not apply this analysis to English *only*, but to German *nur* and Italian *solo*. She suggests that English *just* may be unspecified just like *nur* or *solo*.

Besides issues of parsimony, this account does not naturally extend to French *seulement*, which does not combine with an additive particle in negative contexts to give rise to a most-likely reading. More problematically, French *seulement* exhibits two readings under negation as was shown in (36)–(37), which is not predicted by this analysis. For these reasons, while adopting the scalar analysis of *only*, I propose a different analysis for *seulement* and scalar particles in general, which I lay out in the next section.

4. A new theory of scalar particles

4.1 French *seulement*

To account for the distribution of readings shown by *seulement* (two in negative contexts, one in positive contexts), I propose to analyze this particle as a conjunction and derive the double reading under negation from the interaction between conjunction and negation. Specifically, I first assume that *seulement* is always scalar (cf. Klinedinst 2004) in implying that the alternatives are less likely. The apparent difference between standard *seulement* and scalar *seulement* comes from the nature of the scale; *seulement* seems to exclude either lower alternatives on an expectedness scale (and thus higher on a significant scale) or higher alternatives on a numerical scale. But this is logically equivalent as higher alternatives on a numerical scale are less likely than lower alternatives (since the former logically imply the latter). This is illustrated in (43).[12]

(43) a. *Paul a seulement une licence.*
Paul has *seulement* a license
'Paul only has a BA.'

b. *Paul a seulement un diplôme universitaire.*
Paul has *seulement* one diploma academic
'Paul only has one degree.'

12. This means that a sentence like (vii) has two readings: if the relevant scale is the numerical one, *only* conveys the meaning that John did not see other people than Mary; if the relevant scale is that of significance/expectedness, (vii) implies that Mary was the most insignificant/likely person for John to see.

(vii) *John only saw Mary.*

Note that the assumption of a unified *only* is not crucial to the proposal though, which is compatible with the hypothesis of an ambiguity between scalar *only* and non-scalar *only*.

(43a) conveys the idea that having a BA is more likely, and thus less significant, than having a higher diploma; similarly, (43b) implies that having one academic degree is more likely, and thus less significant, than having more than one. Furthermore, these less likely alternatives are excluded in both cases. In short, *seulement* consists of the following meaning components.

(44) *Seulement* (p) (first version)
Exclusivity: $\forall q \in C\ [q \neq p \rightarrow q(w) = 0]$
Scalarity: $\forall q \in C\ [q \neq p \rightarrow p >_{\text{likely/insignificant}} q]$
Factivity: $p(w) = 1$

This is basically Guerzoni's (2003) lexical entry of *only*, except that I didn't specify whether these meaning components are presuppositional or assertive. Instead of doing so, I propose to conjoin these meaning ingredients as in (45).

(45) *Seulement* (p) (final version)
$\forall q \in C / q \neq p$, *seulement* (p) = $p \wedge \neg q / p > q$

In words, *seulement* has a conjunctive meaning: it asserts that the proposition it takes as argument is true and that the alternatives lower on a likelihood scale are false. For instance, (46) means that Luke came once and he didn't come more than once.

(46) *Luc est seulement venu une fois*
Luke is *seulement* come one time
'Luke only came once.'

Crucially, the hypothesis of a conjunctive meaning correctly predicts a disjunctive meaning in negative contexts – assuming that *seulement* scopes under negation – because of De Morgan's law as explained in (47) and illustrated in (36) repeated in (48).

(47) a. De Morgan's law: $\neg (p \wedge q) \leftrightarrow \neg p \vee \neg q$
b. not (*seulement* (p)) $= \neg (p \wedge \neg q)$
$= \neg p \vee q$

(48) *Luc n' est pas seulement venu une fois*
Luke NEG is not *seulement* come one time
Only-Meaning: Luke didn't only come once.
Even-Meaning: Luke didn't even come once.

Based on (47b), (48) means that Luke did not come once or he came more than once. The members of the disjunction precisely correspond to the *even*-like and the *only*-like meanings respectively.

4.2 French *même, quand même, ne serait-ce que*

The meaning of the other French scalar particles mentioned in the previous sections can be analyzed using the same parameters:

a. conjunction/disjunction of alternatives
b. true/false alternatives
c. more/less likely alternatives
d. scope under/over negation

Thus, *même,* just like *seulement,* expresses a conjunction, but it differs from it with respect to the three other factors as shown in (49). That's why *même* can induce the same reading as *seulement* in negative contexts, despite different readings in positive environments.

(49) *même* (p) (final version)
$\forall q \in C/q \neq p$, *même* (p) = $p \wedge q$ / $p < q$

Quand même only differs from *seulement* with respect to its scope with the negation. That's why they display different readings only in anti-additive contexts. The meaning in (50) also captures the antonymy with *même*: *quand même* and *même* both have a conjunctive meaning and scope over negation, but they differ with respect to the two parameters concerning the alternatives.

(50) *quand même* (p) (final version)
$\forall q \in C/q \neq p$, *quand même* (p) = $p \wedge \neg q$ / $p > q$

As for *ne serait-ce que*, which we saw is close in meaning to 'at least', I propose to analyze it as a disjunction: it conveys the meaning that the proposition it takes as argument or less likely alternatives are true.

(51) *ne serait-ce que* (p) (final version)
$\forall q \in C/q \neq p$, *ne serait-ce que* (p) = $p \vee q$ / $p > q$

This correctly predicts the equivalence in meaning between *même* and *ne serait-ce que* in negative contexts, given that *ne serait-ce que* scopes under negation, as opposed to *même*. This is shown in (52) and (53).

(52) a. De Morgan's law: $\neg (p \vee q) \leftrightarrow \neg p \wedge \neg q$
b. $\neg$ (*ne serait-ce que* (p)) $= \neg (p \vee q)$
$= \neg p \wedge \neg q$

(53) *Luc n' est (même) pas venu (ne serait-ce qu') une fois.*
Luke NEG is *même* not come *ne serait-ce que* a time
not (***ne serait-ce que*** (Luke came once))
même (Luke did **not** come once)
= Luke did not come once and he did not come more than once

Table 6 summarizes the analysis of the French scalar particles examined in this paper. This correctly predicts their distribution according to their meaning as represented in Table 7.

Table 6. Analysis of French scalar particles

	Meaning	Position on the likelihood scale	Scope with negation
seulement	$p \wedge \neg q$	$q < p$	NEG > *seulement*
ne serait-ce que	$p \vee q$	$q < p$	NEG > *ne serait-ce que*
quand même	$p \wedge \neg q$	$q < p$	*quand même* > NEG
même	$p \wedge q$	$q > p$	*même* > NEG

Table 7. Distribution and interpretation of French scalar particles

Position on scale of focus associate/Context of occurrence	**Positive**	**Anti-additive**	**Other DE**	**Modal, non monotone**
Top e.g. 'invite the pope'	*même*	*quand même*	*même*	*même*
Bottom e.g. 'invite one's best friend'	*seulement-only* *quand même*	*même* *ne serait-ce que* *seulement-only* *seulement-even*	*ne serait-ce que* *seulement-only* *seulement-even* *quand même*	*ne serait-ce que* *seulement-only* *quand même*

In sum, building the additivity, exclusivity and scalarity of scalar particles into a conjunctive or disjunctive meaning and determining their scope with respect to the negation allow us to parsimoniously derive the different meanings and distributions of French scalar particles.

4.3 Crosslinguistic consequences

This theory predicts the existence of various scalar particles depending on the parameters (a)–(d) mentioned above. For instance, the particle defined in (54), which only differs from *seulement* with respect to the scalarity component, is expected to exist.

(54) $\forall q \in C/q \neq p$, *particle* $(p) = p \wedge \neg q$ / $p < q$

Based on Tomaszewicz's (2012) study, this prediction appears to be borne out: slavic *aż/čak* can be defined as in (54), as illustrated in (55) for Polish.

(55) *Maria jest aż menedżerem*
Maria is *aż* manager
'Maria got as far as being the manager.'

It would be worth testing more of these predictions based on crosslinguistic work.

What about English? Table 8 summarizes the distribution of (scalar) *only* and *even* depending on their meaning.

Table 8. Distribution and interpretation of English *only* and *even*

Position on scale of focus associate/Context of occurrence	Positive	Anti-additive	Other DE	Modal, non monotone
Top e.g. 'invite the pope'	least-likely *even*		least-likely *even*	least-likely *even*
Bottom e.g. 'invite one's best friend'	*only*	most-likely *even* *only*	most-likely *even* *only*	most-likely *even* *only*

Thus, English *even* seems to correspond to both French *même* ($p \wedge q$) and *ne serait-ce que* ($p \vee q$). Tentatively, I hypothesize that *even* is underspecified between these two meanings and the stronger meaning obtains in the context, namely, $p \wedge q$ in positive environments and $p \vee q$ in negative environments.

Conversely, English *only* is more restricted than French *seulement*: it only exhibits one reading under negation. One possibility is thus to assume that, similarly, *only* is underspecified between $p \wedge \neg q$ and $p \vee \neg q$. This should be examined in future work.

5. Conclusion

In this paper, I have provided new empirical observations about French scalar particles, which behave differently from English scalar particles despite what is usually assumed. In particular, French *même* presents a more restricted distribution than English *even* and is complemented by *ne serait-ce que*, which provides further arguments for the scope theory against the ambiguity theory of *even*-like particles. Conversely, the distribution of French *seulement* is less restricted than that of English *only*, which reveals the existence of an intrinsic link between *even*-like particles and *only*-like particles. To capture these new facts and more generally

derive the organic relations between scalar particles, I have built a new theory of scalar particles, the behavior of which is claimed to depend on the following parameters: conjunctive/disjunctive meaning; true/false alternatives; more/less likely alternatives; scope under/over negation. Thereby, we do not face the problems of previous accounts, which have to assume island violation or non-economical mechanisms such as multiple ambiguity or assertion/presupposition swapping. This theory makes many crosslinguistic predictions that should be tested in the future.

References

Beaver, David I., and Brady Z. Clark. 2008. *Sense and Sensitivity: How Focus Determines Meaning*. Blackwell Pub. doi:10.1002/9781444304176

Crnič, Luka. 2011. *Getting* even. MIT Dissertation.

Gast, Volker, and Johan van der Auwera. 2011. "Scalar Additive Operators in the Languages of Europe." *Language* 87 (1): 2–54. doi:10.1353/lan.2011.0008

Giannakidou, Anastasia. 2007. "The Landscape of EVEN." *Natural Language and Linguistic Theory* 25 (1): 39–81. doi:10.1007/s11049-006-9006-5

Guerzoni, Elena. 2003. *Why Even Ask? On the Pragmatics of Questions and the Semantics of Answers*. MIT Dissertation.

Herburger, Elena. 2000. *What Counts: Focus and Quantification*, vol. 36. The MIT Press.

Hoeksema, Jack, and Hotze Rullmann. 2001. "Scalarity and Polarity: a Study of Scalar Adverbs as Polarity Items". In *Perspectives on Negation and Polarity*, ed. by Jack Hoeksema *et al.*, 129–171. Amsterdam: John Benjamins. doi:10.1075/la.40.07hoe

Horn, Laurence R. 1971. "Negative Transportation: Unsafe at any Speed". *CLS* 7: 120–133.

Karttunen, Lauri, and Stanley Peters. 1979. "Conventional Implicature." In *Syntax and Semantics*, ed. by C.-K. Oh, and D. A. Dinneen, vol. 11, 1–56. Academic Press.

Klinedinst, Nathan. 2004. "Only Scalar *Only*." Handout at Presupposition and Implicature Workshop, Paris.

König, Ekkehard. 1991. *The Meaning of Focus Particles*. London: Routledge. doi:10.4324/9780203212288

Lahiri, Utpal. 1998. "Focus and Negative Polarity in Hindi." *Natural Language Semantics* 6: 57–123. doi:10.1023/A:1008211808250

Lerner, Jean-Yves, and Thomas E. Zimmermann. 1981. "Mehrdimensionale Semantik. Die Präsuppositionen und die Kontextabhängigkeit von *nur* [Multidimensional Semantics. The Presuppositions and Context Dependency of *nur*]." *Arbeitspapier des Sonderforschungsbereichs 99*. Universität Konstanz

Nakanishi, Kimiko. 2006. "*Even*, *Only*, and Negative Polarity in Japanese." *Proceedings of SALT* 16: 138–155. doi:10.3765/salt.v16i0.2953

Rooth, Mats. 1985. *Association with Focus*. University of Massachusetts, Amherst Dissertation.

Rooth, Mats. 1992. "A Theory of Focus Interpretation." *Natural Language Semantics* 1 (1): 75–116. doi:10.1007/BF02342617

Rullmann, Hotze. 1997. "*Even*, Polarity, and Scope." *Papers in Experimental and Theoretical Linguistics* 4: 40–64.

Schwarz, Bernhard. 2005. "Scalar Additive Particles in Negative Contexts." *Natural Language Semantics* 13: 125–168. doi:10.1007/s11050-004-2441-0

Shank, Scott. 2002. "*Just* and its Negative Polarity Variants in Salish." Ms. University of British Columbia, Vancouver.

Tomaszewicz, Barbara M. 2012. "A Scalar Opposite of Scalar *Only*." *Proceedings of the 30th West Coast Conference on Formal Linguistics*: 324–334.

Wilkinson, Karina. 1996. "The Scope of *Even*." *Natural Language Semantics* 4: 193–215. doi:10.1007/BF00372819

Zwarts, Franz. 1998. "Three Types of Polarity." In *Plurality and Quantification*, ed. by F. Hamm, and E. Hinrichs, vol. 69, 177–238. Kluwer. doi:10.1007/978-94-017-2706-8_5

CHAPTER 4

Pluralities of events

Semelfactives and a case of 'single event' nominalisation

Marta Donazzan and Lucia M. Tovena
University of Cologne & Emerging Group DSLC / Université Paris VII

In this paper, we tackle the issue of the semantics of semelfactive predicates by looking at *ata*-nominalisations built on instrument semelfactive verbs in Italian. Being *nomen vicis* forms, these nominalisations denote singular instances of events. The comparison allows us to cast new light on the issue of what counts as a unit in the domain of semelfactive verbs. We then apply this line of exploration to the issue of the double reading of semelfactives – semel and processive, and show that the two readings of their *ata*-nominalisations correspond to predicates belonging to distinct aspectual types.

Keywords: event plurality, aspect, *ata*-nominalisations, semelfactives

1. Introduction

Verbs like *bastonare* in Italian describe events of hitting where the action is performed with an instrument (1). They admit two interpretations. One reading, called *semel,* is illustrated by sentence (1a) that can be used felicitously in a context where Gianni gave only one blow with a stick to the pot. The predicate may also be interpreted as denoting instances of activity, called *processive,* as sentence (1b) is true if Gianni repeatedly hits the puppet with a stick.

(1) a. *Gianni ha bastonato la pignatta ed è caduta una pioggia di monete*
Gianni has hit the pot and is fallen a rain of coins
'Gianni hit the pot and coins fell down everywhere.'

DOI 10.1075/rllt.11.04don

b. *Gianni ha bastonato il pupazzo che si era ribellato*
Gianni has beaten the puppet that CL was rebelled
'Gianni beat the puppet who had rebelled.'

One blow is a quantum denoted by semelfactive *bastonare*. Blows in a series denoted by processive *bastonare* are identical, for all practical purposes. It is a matter of debate whether one reading is derived from the other, i.e. whether the processive reading is a pluralisation of singular blows. The issue is made more complex by the fact that in Romance, as in English, the verbal form conveying the two readings is the same. And yet, there seems to be more than mere pluralisation in the case of the processive reading. The presence of an expression that can convey information on 'why' an 'action' has been performed seems to ease the interpretation of the verb as describing a single action. The modifier *who had rebelled* in (1b) enhances the accessibility of the processive reading for *bastonare*.

In this paper, we contribute to the discussion by examining a type of nominalisation in Italian that denotes singular instances of events and receives an interpretation that has similarities to that of semelfactive verbs. 'Delimitedness' of events need not be understood in terms of telicity. In a number of Romance languages, a speaker can use *nomina vicis* forms to make reference to single instances of contingently delimited events. Nomina vicis are expressed by event nouns formed with the suffix *-ata/-ada* (see Gaeta 2000; Aliquot-Suengas & Macchi 2003; Acquaviva 2005, among others), that contribute a type of delimitation that is better described as aspectual boundedness. In the remainder of this section, we compare *ata*-nouns *nomina vicis* and semelfactive predicates.

1.1 Semelfactives and *ata*-nominalisations

In Romance languages, semelfactives are a class of verbs denoting singular occurrences of events, which are also systematically coupled with atelic activity predicates, see *bussare* 'knock' (2).

(2) *Mario ha bussato alla porta. (= un colpo/diversi colpi)*
Mario has knocked at-the door. (= one knock/several knocks)

Semelfactive *tossire* 'cough' (3), when modified by *time* adverbials, denotes the exact number of single coughs by Gianni (3a). Durative modifiers (3b) enhance its processive interpretation.

(3) *Gianni ha tossito*
'Gianni coughed
a. …*una volta, e il colpo di tosse ci ha svegliato*
…once, and the cough woke us up.'
b. …*per cinque minuti*
…for five minutes.'

There is no agreement in the literature on whether semelfactives are an aspectual class (Smith 1991) or not (Dowty 1979). In some languages, for example in the Slavic group, affixation marks forms that do not belong to a unique aspectual class.[1] Since we are discussing a Romance language here, we follow Smith (1991) and consider semelfactive predicates to be an aspectual class. The defining criteria of this class are one of the two important issues at the heart of the debate on semelfactive verbs. Smith (1991) characterises them as denoting non-durative, non-telic dynamic events. These events are delimited without culmination and change, which means that, contrary to achievements and accomplishments predicates, semelfactives do not have a well defined endpoint, do not imply the presence of a resultant state, nor a transition point to a potential resultant state. This type of delimitedness is not further explained in Smith's theory of lexical aspect, since semelfactives are neither properly atelic nor strictly telic. Alternative solutions are offered by Rothstein (2008), who characterises the events denoted by semelfactives as naturally atomic individuals (hence telic events) that can be construed into sums, and by Tovena (2010b), according to whom the events described by semelfactive verbs are characterised by cycles of parts. The execution of at least one full cycle makes possible the realisation of the event as a minimal instance – not necessarily unstructured and with no standard beginning.

Empirical evidence pointing towards delimiteness without telicity is provided by the fact that semelfactives in their semel reading do not seem to be coercible in the way predicates denoting instantaneous telic events are. Recall that change-of-state predicates like achievements can be coerced into denoting an activity that is understood as the "preparatory phase" of the event, for instance when they are modified by imperfective aspect, cf. the achievement *arrivare* 'arrive' modified by the progressive periphrasis in (4).

(4) *Mario stava arrivando in stazione (quando ha bucato la ruota)*
'Mario was about to arrive at the train station (when he punctured the tyre).'

1. For a more detailed discussion on the origins of semelfactivity in Slavic and other languages, see Nesset (2013). Nesset shows that the property of instantaneousness typical of Western Europe semelfactive verbs characterises the oldest instances of semelfactive verbs in Russian too, although it is no longer a requirement nowadays.

However, the progressive cannot coerce semelfactives in the same way; see Example (5). This is coherent with the proposal that semel predicates do not denote a change of state (Smith 1991; Tovena 2010b), contrary to achievements.

(5) *Mario stava bussando quando ho aperto la porta (#e ha mutato il colpo in un saluto)*
'Mario was knocking when I opened the door, (#and turned the knock into a hand waving).'

Bussare is interpreted as an ongoing activity in (5), and the proposition expressed by the sentence is true in the standard conditions for progressive, namely in a situation where Gianni has already given at least one knock (he has already started knocking) when I open the door. The sentence with the progressive periphrasis entails a sentence with the *passato prossimo*/perfect as per Vendler's test. The continuation in parentheses provides a description of an event subsequent to the door opening and cannot be taken to provide a different characterisation of the described event, which would be understood as the preparatory phase of an arguably telic event of knocking. Romance languages distinguish the progressive interpretation from the inceptive one. Only the inceptive periphrasis can focus on the preparatory phase of the event (6a), as it is the case also for uncontroversial activity predicates (6b). In both cases in (6), the continuation is interpreted as providing a different description of the same act.

(6) a. *Mario stava per bussare quando ho aperto la porta (e ha mutato il colpo in un saluto)*
'Mario was about to knock when I opened the door (then he turned the knocking into a hand waving).'

b. *Mario stava per parlare quando ho aperto la porta (e ha fatto finta di sbadigliare)*
'Mario was about to speak when I opened the door (and he pretended to be yawning).'

A second issue discussed in the literature on semelfactives, is how to represent the systematic ambiguity of these predicates in some languages and whether one of the two readings of (1) and (3) is derived from the other. Rothstein (2008) describes semelfactive verbs as homophonous with atelic activity predicates, but is not explicit about the conditions under which the cumulative summative operation that justifies this systematic relation can apply. According to Tovena (2010b), the processive reading corresponds to instances where the cycle characterising the event is executed more than once. We adopt her notion of cycle that allows us to carve out

units of event without assuming the presence of a termination point or atomicity. Thus, the two readings are independent but related.

Turning to *ata*-nominalisations, the contribution of lexical and grammatical aspect to the delimitation of the events they denote and their exact characterisation as singular units is still unsettled (see Ippolito 1999; Gaeta 2000; Acquaviva 2005, and Section 2 below). On the one hand, the events denoted by *ata*-nominalisations are not intrinsically (near)instantaneous like semelfactives, e.g. a *nuotata* can be short or have a certain duration (7). This empirical difference makes Tovena (2014) claim, contra Gaeta (2000), that the denotation of *ata*-nouns as *nomina vicis* (see Section 2.2) must be kept distinct from that of semelfactive predicates.

(7) *Ha fatto una nuotata di diverse ore*
has done a swim-ATA of several hours
'He went for a swim lasting several hours.'

On the other hand, there are similarities because both Italian *ata*-nominalisations and semelfactives have entities that count as singular units in their domain. In the case of *ata*-nouns, boundedness is not inherited by the class of the base verb. Yet, some lexical restrictions govern the well-formedness of the nominalisations.

1.2 Structure of the paper

In this paper we compare predicates that exibit semelfactive and processive readings to *ata*-nominalisations built on them, in order to gain insights on the semantics of these items and shed new light on the notion of plurality and unity of events. Section 2 introduces *ata*-nominalisations and focuses on their *nomen vicis* interpretation. We look at the aspectual constraints that trim the set of verbal bases entering *ata*-nominalisations, and compare them to the constraints that govern derivations from nominal bases. Next we consider derivations where the base can either be an instrumental semelfactive verb or the noun that denotes the relevant instrument. In Section 3, we cash in on this distinction and show that these constraints are best accounted for adopting the hypothesis that these event nominals are construed following two distinct derivation patterns. This leads us to conclude that instrument semelfactives, in their processive reading, are instances of activity predicates, which belong to a distinct aspectual type and cannot be considered as the mere pluralisation of semel events.

2. *ata*-nominalisations and delimited events

2.1 Origin of the form

Nominalisations ending in *-ata* are derived from nominal and verbal bases.[2] Deverbal nominalisations can be analysed diachronically as obtained by adding the feminine ending *-a* to the past participle form (Ippolito 1999). Evidence for this claim would come from nominalisations of irregular participial forms (such as *corsa* from *correre* 'run', irregular past participle form *corso*), or nominalisations of verbs of the 2nd/3rd conjugation, whose thematic vowel is *-u/i*, cf. *bevuta* from *bere* 'drink' and *dormita*, from *dormire* 'sleep'.

Under this view, the diachronic origin of the suffix *-ata* would explain the aspectual properties of *ata*-nouns, since the perfective feature of the participial suffix would contribute the bounded event interpretation and the tokenisation effect of the nominalising suffix. However, there are at least two arguments pointing towards the conclusion that the *-ata* participial ending has gained autonomy and is now an independent suffix with its own semantic content. Synchronically, the suffix can attach also to nominal bases, yielding a name of event in this case too. We introduce these cases in Section 2.2. Second, past participle formation is not sensitive to the aspectual class of the base verb in the way *ata*-nominalisations are, as noted by Tovena (2014). This sensitivity to the aspectual properties of the base is discussed in Section 2.3, where we look at the output of the nominalisation of different types of atelic predicates and see that the aspectual properties of the base constrain the productivity of the nominalisation. These facts lead us to assume that *ata* has become synchronically an independent derivational suffix with specific properties.

2.2 Delimited events and the reading *nomen vicis*

There is no formal definition of the semantic class of *nomen vicis* in the literature. We use this term to characterise event nouns that are specialised for a token interpretation. Romance nominalisations derived with the suffix *-ata/ada* are *nomen*

2. Since in this paper we are interested in event nouns, we do not consider the patterns in (i)–(iii) below.

(i) *cucchiaio* (spoon) – *cucchiaiata* – the quantity of *x* carried by N
(ii) *notte* (night) – *nottata* – the time span covered by N
(iii) *peperone* (pepper) – *peperonata* – food obtained by grinding/smashing N

Furthermore, we do not discuss the case where the output denotes an act of prototypical behaviour (e.g. *ragazzo*, boy – *ragazzata*, childish act). For a broader survey see Gatti and Togni (1991); Acquaviva (2005); von Heusinger (2002), and references therein.

vicis forms in this sense, as they specifically denote singular events in a count domain, see (8), where *nuotata* refers to a specific event of swimming.

(8) *Gianni ha fatto una nuotata*
Gianni has done a swim-ATA
'Gianni had a swim.'

A characterisation of the reading *nomen vicis,* at least in negative terms, is that they do not admit non-token, generic interpretations (Gaeta 2000) cf. the unacceptability of generic statements where *ata*-nominalisations are headed by definite singular DPs and are interpreted as referring to the kind, cf. (9). Generic interpretation of noun phrases in combination with individual predicates has been analysed as involving realisations of kinds (Carlson 1977). Kinds are the maximal sum of individuals in the denotation of the noun with respect to a world. Such reinterpretation of the maximal sum does not seem possible for a *nomen vicis* item.

(9) **La nuotata rilassa i muscoli* (Gaeta 2000)
the swim-ATA relaxes the muscles

There are seemingly counterexamples to this generalisation. A reviewer notes that the nominalisation *ammucchiata*, which s/he sees as derived from *ammucchiare*, 'lump together', admits a generic reading (10).

(10) *L'ammucchiata è passata di moda*
the-ammucchiata is passed of fashion
'Orgies are out of fashion.'

The counterexample is not a real one, since the generic reading of the DP is not obtained by simple type shifting of the corresponding event noun, as the English translation makes clear, but requires some semantic enrichment. The act of lumping here refers specifically to a gathering of people and denotes a type of sexual practice. Under a less specific interpretation, the event noun is infelicitous, also in a token interpretation (11a) and even when it denotes established social practices as in (11b).

(11) a. #*Ho fatto un'ammucchiata di roba vecchia da buttare*
have made a-lump-ATA of stuff old to throw
'I lumped together old stuff to be thrown away.'

b. ??*La sciata a Cortina è passata di moda.*
the ski-ATA in Cortina is passed of fashion
'Skiing in Cortina is out of fashion.'

Second, the generic interpretation of event *ata*-nouns (12) requires plural marking, which is not generally required for nouns in Italian, be they event nouns or not (13).

(12) *Le camminate fanno bene alla salute*
the walk-ATA-PL do good to-the health
'Hiking is good for one's health.'

(13) a. *Il nuoto rilassa i muscoli*
'Swimming relaxes the muscles.' (Gaeta 2000)
b. *La scollatura vertiginosa è di moda.*
'A plunging neckline is fashionable.'

Third, the delimitedness of the events denoted by *ata*-nouns is an intrinsic property that is not inherited from the *Aktionsart* of the verbal base. The event is presented as bounded even if the base verb is atelic, as in the case of *nuotare* 'swim' in (8). More to the point, *ata*-nominalisations do not support the construction of cumulative events, cf. (14) (Tovena 2014). This property distinguishes *nomina vicis* from instances of activities, whose inherent cumulativity follows from their being atelic predicates (Dowty 1979).

(14) If I made a *nuotata* from 9am to 10am and a *nuotata* from 10am to 11am, it does not follow that I made a *nuotata* from 9am to 11am.

Finally, *ata*-nominalisations can yield event nouns also in the absence of a verbal base, and the denoted events have the same aspectual properties, cf. (15) and (16), for which putative base verbs °*ombrellare*[3] and °*padellare* are not attested (Scalise 1984; Acquaviva 2005, a.o.).

(15) *Mario ha dato un'ombrellata a Luca.*
Mario has given a-umbrella-ATA to Luca
'Mario hit Luca once with an umbrella.'

(16) *Gianni ha dato una padellata in testa a Luca*
Gianni has given a pan-ATA on head to Luca
'Gianni hit Luca on the head with a frying pan.'

The event nouns built on a nominal base have a *nomen vicis* interpretation. They denote singular events of hitting with an instrument, and the bases characterise the instruments. Only the semel reading appears to be accessible, cf. (17) vs. (23) below.[4]

3. The sign '°' marks non-existent forms.

4. See footnote 2 for other examples of *ata*-nominalisations built on nominal bases, which we will not discuss in this paper.

(17) *Gianni ha dato un'ombrellata a Luca*
Gianni has given a umbrella-ATA to Luca

a. ^{OK} un colpo (a blow)
b. #diversi colpi (several blows)

In other words, *ombrellata* cannot be interpreted as an instance of the activity of beating with an umbrella for a delimited stretch of time, but can only be understood as referring to an event of hitting once with this instrument.

Since we are interested here in comparing *ata*-nouns and semelfactive verbs, which are ambiguous between a singular vs. plural reading, in the remainder of this section we focus on nominalisations built on verbal bases, which may exhibit the same ambiguity. We come back to this observation concerning nominal bases that do not admit processive interpretations in Section 3.

2.3 Verbal bases and aspectual constraints

The productivity of *ata*-nominalisations built on verbal bases is constrained by the aspectual properties of the base verb insofar as only atelic dynamic predicates can provide verbal bases. States and accomplishments are excluded, and achievements must be coerced into extended events (Gaeta 2000; Tovena 2014). Compare (8) and (18).

(18) a. **Gianni ha fatto un'arrivata*
Gianni has done an-arrive-ATA

b. **Gianni ha fatto una costruita della casa*
Gianni has done a build-ATA of-the house

As for the contrast between (18a) and (19a), note that in (19a) the modification of the N suggests that the achievement predicate *entrare* 'enter' is interpreted as a complex event; the entering possibly includes cheering, head-turning, and hand-clapping. In (19c), the processive reading of the unmodified predicate is possible in virtue of the subject denoting an individual with spatial extension, allowing the event to be stretched over an extended temporal interval, an option not available for (19b).[5]

5. As an aside, we discuss some potential counterexamples correctly pointed out by one reviewer. The nominalisations of *entrata* and *uscita* are acceptable in sentences such as (i), where the NP either does not denote properly in the event domain (cf. *entrata* as way-in in (ia)) and, therefore, is not relevant to our argument, or it acquires the 'extended event' reading observed for (19a, c). (i) could be uttered by an actor referring to the opening of his performance, and it is, thus, implicitly understood as a more complex event with some duration (ib).

(19) a. *Gianni ha fatto un'entrata trionfale*
Gianni has done an-enter-ATA triumphant
'Gianni made a triumphant entry.'

b. #*L'entrata di Gianni in stazione*
the-enter-ATA of Gianni in station
'The entering of Gianni into the train station.'

c. *L'entrata del treno in stazione*
the-enter-ATA of-the train in station
'The entering of the train into the train station.'

Next, transitive predicates can be nominalised *via* the *-ata* suffix only if their internal complement is interpreted non-referentially, cf. *mangiare* 'eat' in (20b). Recall that the notion of incremental theme has been used precisely to refer to the dependence between quantitative information in nominals and (a)telicity in events. The accomplishment in (20a) is acceptable only when 'detelicised' by the bare noun complement of the preposition in the PP *di funghi* 'of mushrooms' (20b) (Donazzan & Gritti 2013).

(20) a. **Gianni ha fatto una mangiata dei funghi*
Gianni has done an eat-ATA of-the mushrooms

b. *Gianni ha fatto una mangiata di funghi*
Gianni has done an eat-ATA of mushrooms
'Gianni had a mushroom treat.'

(i) *Ho sbagliato l'entrata.*
have mistaken the-entrata
a. I took the wrong way in.
b. I messed up the entrance.

Next, the same reviewer questions the ban on telic predicates pointing out the case of the nominalisation *rimpatriata*, which, s/he argues, derives from the verb *rimpatriare* 'come/go back to the homeland', itself denoting a telic event. While such a derivation is plausible, synchronically, *rimpatriata* appears to have only a specialised interpretation denoting a gathering of old acquaintances, cf. (ii). Results from a Google search, and judgments of native speakers we consulted, who consistently reject sentences such as (iii), where the 'back to homeland' reading is enhanced, strengthen our claim.

(ii) *Ho fatto una rimpatriata in pizzeria con gli amici*
have made a rimpatriata in pizza-restaurant with the friends
'I met with my old friends to have a pizza together.'

(iii) #*Ho fatto una rimpatriata a Natale per rifare il passaporto.*
have made a rimpatriata at Christmas to renew the passport
'I went back home at Christmas for renewing the passport.'

The ban on definiteness is to be interpreted in aspectual terms. Objects that do not measure out the event can be definite (Tovena 2014), as illustrated by the change-of-state predicate *pulire* ('clean') in (21), which is independently ambiguous between telic and atelic interpretations (Hay et al. 1999; Kearns 2007).

(21) *Ha dato una pulita alla stanza*
had given a clean-ATA to-the room
'S/he cleaned the room (a bit).'

The ban on telic readings could be put to use to explain the interpretation of *ata*-nouns built on semelfactive verbs. When the base is a semelfactive verb, such as *tossire* 'cough' (22)[6] or *bussare* 'knock' (23), the nominalisation selects the processive interpretation and denotes a delimited instance of an activity.

(22) *Mario ha fatto una tossita*
Mario has made a cough-ATA
'Mario coughed/gave a coughing.' (www)

The semel reading (a) is hardly accessible in (23).

(23) *Gianni ha dato una bussata al finestrino per vedere se tutto andava bene*
'Gianni gave a knocking at the window to see if everything was fine.'
a. ??*un colpo* (a knock)
b. OK *diversi colpi* (several knocks)

The data in (22) and (23) fit the predictions of the aspectual constraint that governs the productivity of *ata*-nominalisations, because the predicate is interpreted as properly atelic only in the activity reading.

Internal pluractional predicates such as *saltellare* 'leap about' are another case where *ata*-nominalisations seem to inherit their homogeneous internal structure from the base (Tovena 2010a). The event denoted by the nominalisation is a single event made up by a plurality of Cusic's (1981) phases, sub-events that are not accessible to counting and distributive effects. See (24).

(24) a. *Gianni è sceso a valle facendo una saltellata giù per il ghiaione*
Gianni has descended to valley doing a gambol-ATA down across the scree
'Gianni reached the valley gambolling down the scree.'
b. #*Alla quinta saltellata si è storto una caviglia*
'At the fifth leap he sprained his ankle.'

6. The degree of acceptability of the nominalisation *tossita* varies across speakers.

The empirical situation discussed so far can be summarised as follows. All verbs that admit *ata*-nominalisations describe events with some duration. In the case of deverbal *ata*-nominalisations built on activities, the events denoted by the nominalisations and their verbal bases express similar event description *modulo,* the addition of boundedness. *Ata*-nominalisations built on pluractional or semelfactive predicates yield instances of internally pluractional activities. Finally, in the case of denominal forms, *ata*-nominalisations display semel readings and denote events of 'hitting with an instrument' where the bases characterise the instruments.

2.4 The case of instrument semelfactives

The observation that denominal *ata*-nominalisations only have semel readings is at first contradicted by *bastonata* in (25), which supports two readings, and in this respect is similar to what we saw in (1) for the semelfactive verb.

(25) *Gianni ha dato una bastonata al fantoccio*
Gianni has given a stick-ATA to-the puppet

The semel reading of *bastonata* is made more perspicuous by the context described in (26), and the processive reading is illustrated by (27), where the event can be described as an instance of beating with a stick.

(26) *Mario ha dato una bastonata al fantoccio, e il colpo ha*
Mario has given a stick-ATA to-the puppet, and the blow has
tranciato di netto la gamba
cut of clean the leg
'Mario hit the puppet with a stick, and the blow cut the leg clean off.'

(27) *Mario ha dato una bastonata al fantoccio, ma dopo qualche*
Mario has given a stick-ATA to-the puppet, but after some
colpo il bastone si è rotto
blow the stick CL is broken
'Mario hit the puppet with a stick but after a few blows the stick broke.'

The peculiarity of nominalisations like *bastonata*, however, is also that two bases can be envisaged, a verbal and a nominal one. Each option by itself is in agreement with what is said for either denominal or deverbal *ata*-nominalisations, and the two taken together provide the full coverage. The reading as instance of activity would be paired with a verbal base, e.g. the verb *bastonare* in (1), and the semel reading would pertain to *ata*-nominalisations built on a nominal base, the noun *bastone* 'stick'.

In the next section, we explore the possibility of assuming a double base. The discussion leads us also to examine cases like *ombrellata*, for which a verbal base °*ombrellare* cannot be posited, and explain why they cannot be interpreted as instances of an activity of beating with an instrument.

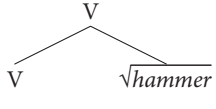
3. Bases and readings

3.1 The double derivation hypothesis

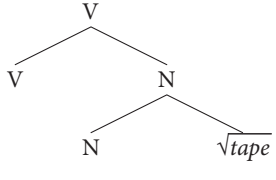
The hypothesis of expressing the opposition between nominal and verbal bases as a distinction in the derivation of an event noun is to be read in the broader perspective of distinguishing between creating words from roots and creating words from existing words. We are going to exploit it, in particular, to make correct predictions with respect to the aspectual interpretation of the derived *ata*-nouns.

Arad (2003) ascribes to Marantz (2000) the idea of bringing into Distributed Morphology the traditional distinction between creating words from roots and from existing words, and applies it to Modern Hebrew, a language that typologically would have only a root-based system. The root vs. word distinction, in the technical sense of these terms, is used in Distributed Morphology to reproduce the distinction lexical/derivational formation vs. syntactic/inflectional formation. Formation from roots is 'lower' and may exhibit idiosyncrasies. Within this domain, the interpretation of the combination of root and word-creating head can be multiple, i.e. polysemy is more easily found. On the contrary, the range of interpretations for word-derived forms is tightly related to the meaning of the words from which they are derived. 'Noun-derived verbs are shown here to depend in their interpretation on the noun from which they are derived, while root-derived verbs may take on multiple, semantically various interpretations' (Arad 2003: 740). The structure Arad proposes for root-derived verbs such as *hammer* is reproduced in (28a) and (28b) reproduces the structure for the noun-derived verb *tape*.

(28) a.



b.



This distinction is useful to deal with the difference in the derivation of *ata*-nominalisations. Before discussing the nominalisation, let's look at each of the possible bases, and take the case of *bastone* 'stick' and *bastonare* 'beat with a stick' as an example.

Suppose Italian has a root like 'baston'. Roots that merge with verb category heads can create verbs with a variety of meanings, in the sense that there is some room for the way in which the root can contribute to the overall meaning. For example, the verb *bastonare* means to beat typically, but not exclusively, with a stick. Any other long, rigid and rather slender object could be used as an instrument to perform the action described as *bastonare*, cf. (29).

(29) *Gianni ha bastonato l'asino con un ramo*
'Gianni beated the donkey with a branch.'

On the other hand, any other action performed with a stick could not be called *bastonare*, e.g. pushing something (30).

(30) #*Gianni ha bastonato la palla per farla avanzare*
(Intended: 'Gianni pushed the ball with a stick to make it go forward.')

The verb defines the 'quality' of the action more rigidly than the nature of the instrument used to perform it. As noted by Kiparsky (1982) for English, the root merely names the most typical instrument used for the activity.

Next, the same root may enter the derivation of the noun *bastone*. Roots that merge with noun category heads can create nouns with a potentially equally large variety of meanings (not so for the nominalisations built on them). For example, *bastone* means a wood stick with some thickness and some length that is not associated with a specific function, e.g. it can be used for reaching something far, pocking or hitting.

3.2 Instrument presuppositions

Arad (2003) exploits Kiparsky's observation in her discussion of denominal verbs and her proposal is subsequently taken up by Scher (2006), who tackles a similar case of instrument presupposition for *ada*-nominalisations in Brazilian Portuguese.

Scher (2006) opts in favour of low verbisation for (31a), where *esfaqueada* is derived from *esfaquear* 'stab' and not from the noun *faca* 'knife', and low nominalisation for (31b), where *parafusada* is derived from the noun *parafuso* 'screw'. The infelicity of (31b) shows that the screw must be the instrument used to perform the event, which provides support for the hypothesis of a nominalisation built on a nominalised root.

(31) a. *O João deu uma esfaqueada no ladrão com um punhal velho*
'João stabbed the thief with an old dagger.'
b. #*O João deu uma parafusada no pé da mesa com um prego*
#'João screwed the leg of the table with a spike.'

Instrument presuppositions are found in Italian *ata*-nominalisations too, cf. (32a vs. b).

(32) a. #*Daniele ha dato un'ombrellata all'asino col bastone*
Daniele has given a-umbrella-ATA to-the-donkey with-the stick
b. *Daniele ha dato una bastonata all'asino con l'ombrello*
Daniele has given a stick-ATA to-the-donkey with the-umbrella
'Daniele beat the donkey with the umbrella.'

We propose to apply the Distributed Morphology analysis to *ata*-nominalisations and assume low verbisation of the root as well as low nominalisation for the form *bastonata*.

3.3 Derivation, aspect, and non-existing verbs

The hypothesis of double derivation can be put to further use to explain the aspectual difference of the two readings illustrated in (26) and (27) with *bastonata*. The data in (27) support the option of verbisation of the root. When the verbal projection categorises the root, the meaning component of this constituent is close to the meaning of the verb for the relevant part. *Bastonata* in its processive reading is the atelic predicate that is required for the nominalisation. As a consequence, the nominalisation gets the 'beating' interpretation.

Conversely, the data in (26) provide support for a nominalisation built on a nominalised root. The interpretation is analogous to that of denominal *ombrellata*. When the root nominalises, we assume that there is a verbal projection inside the structure of the nominalisation, in agreement with the tenets of Distributed Morphology and that it is the nominalising suffix *-ata* itself that introduces such an abstract projection contributing dynamic meaning. The verbal projection takes on the semantics of the categorised noun. This yields an abstract event predicate characterising events of using the object as an instrument, which corresponds to the semel reading.

The hypothesis of positing an abstract verb is not new to the literature on *ata*-nominalisations. According to Samek-Lodovici (2003), there always is a verb formation step in the syntactic derivation of *ata*-nominalisations, which need not correspond to an actual word. However, Gaeta (2000) has objected to this analysis

and argued that putative verbs such as °*ombrellare* are cases of back-formation at best. Consider the case of *bottigliata* 'blow with a bottle' in (33).

(33) *Colpito con una bottigliata in faccia, 40enne finisce in ospedale*
'Hit with a bottle on the face, 40-year-old ends up at the hospital.' (www)

Italian has no verb °*bottigliare* that can be the base of *bottigliata*,[7] although from the morphological point of view, the form °*bottigliare* looks plausible. Its semantics, can be understood (if ever) as 'giving blows with a bottle', via a sort of back-formation process from the *ata*-nominalisation. But besides back-formation, the presence of verbs such as *bastonare* offers further plausible support to the pattern 'blow given with x' used for such an interpretation of °*bottigliare*.

We do not wish to make strong commitments on the necessity of an abstract verb in the derivation. Positing a verb projection, however, has the advantage of providing us with a locus for an event template corresponding to a systematic component of meaning, which could be expressed, for instance, in the form of a bundle of features.

Within this frame, we may also venture a hypothesis about the way in which the pattern *hitting with N*, predictably associated to these nominalisations, arises. As noted also by von Heusinger (2002), the productivity of *ata*-nominalisations is sensitive to the semantics of the base noun, and restrictions can be explained by referring to its lexical type. The event reading depends on the possibility of interpreting the noun as an instrument.

Moreover, we observe that in denominal *ata*-nominalisations denoting events of hitting, the potential referent of the base noun supports a limited number of protoagent inferences (Dowty 1991). It has only the ability of causally affecting something and the possibility of moving with respect to a patient. The *hitting with N* reading would arise as a way to actualize the role of instrument of the base. The nominalisation describes events of contact. The contact corresponds to the point where the entity becomes instrument. It does not match with a *telos* that might otherwise induce a change of state. These events are of short duration and, being also non strictly telic, are, comparable in their conceptual structure to the events denoted by semelfactive predicates.

7. There is a verb *imbottigliare* that means to put into bottles, where the action must be a filling and the bottles are containers, but it may be analysed as resulting from parasynthesis and not prefixation.

3.4 Plurality

The hypothesis of a derivational ambiguity characterises the semel vs. processive interpretive difference as an *aspectual* distinction. The two forms share their being derived from the same root, but have no other direct connection, hence their readings are related but independent. The instance of activity in (27) is not the (morphologically unmarked) pluralisation of a singular event in (26).

The explicative power of the interface model assumed so far is possibly weakened when examining the opposition between plural and singular events. First, let's note that denominal and deverbal *ata*-nominalisations can occur in the plural, in which case they denote collections of events. See (34).

(34) a. *Daniele ha dato delle ombrellate*
Daniele has given some umbrella-ATA.PL
'Daniele gave some blows with the umbrella.'

b. *Daniele ha dato delle bastonate*
Daniele has given some stick ATA.PL
(i) 'Daniele gave some blows with a stick.' (plural of semel)
(ii) 'Daniele beat (somebody) several times.'

The interpretation of deverbal *ata*-nominalisations as plural instances of activities (34b.ii) is dispreferred with respect to the reading of a plurality of semel events (34b.i) in this case. However, anticumulativity, which is a property of *nomina vicis*, blocks the reanalysis of these pluralities into singular superevents of the activity type.

Now, we note that not only reanalysis of a plurality of events into an activity is not possible, but that the processive reading of semelfactives is further constrained by the role of the participants. It appears that the distribution of *ata*-nominalisations is not explained in a fully satisfactory way in a model that eschews argument structure by representing syntactic relations in the lexicon. We seem to need some information on the semantic role of the participants in the event.

For the nominalisation to receive a processive reading (25), the predicate seems to require an animated subject, who performs the event in an active way. The same restriction applies to verbs of hitting with an instrument. The activity reading of the verb *frustare* 'whip' is not available with an inanimate agent (35a), which nevertheless can be involved in a plurality of single events of striking denoted by the plural of the nominalisation *frustata*, (35b).

(35) a. #*Il cavo si è rotto e ha frustato il muro fino al*
the cable CL is broken and has lashed the wall until at-the
mattino
morning

b. *Il cavo si è rotto e ha dato delle frustate*
the cable CL has broken and has given some whip-ATA.PL
contro il muro fino al mattino
against the wall until at-the morning
'The cable broke and struck lashes against the wall until dawn.'

What counts for acceptability is the possibility for the subject to satisfy proto-agentive entailments in a more fine-grained way. Besides being animate, an agent can also be *sentient*, i.e. aware of her participation in the event. Sentience is a property of animate entities, but it is not necessarily entailed by animacy. If the agent of the nominalised event is animate but not sentient, only the semel reading remains accessible, cf. (36), where sentience is suspended by a subject-oriented adverbial.

(36) *Mario ha dato una bastonata a Luca senza accorgersene*
Mario has given a stick-ATA to Luca without notice
(#ma dopo qualche colpo il bastone s'è rotto)
(but after some blow the stick CL-is broken
'Mario hit Luca with a stick without noticing (#but after a few blows the stick broke).'

We take the necessity of sentience for the felicity of (36) to reveal to us that a more general principle is at work here, one concerning the perspective under which the event is described. The perspective adopted for the description of an entity has been shown to be relevant for lexical semantics and has been evoked in particular to account for the mass-count or singular-plural distinction in the denotation of nouns. A plural collection considered under a specific perspective can be seen as a whole, i.e. as an entity of a different sort. In the case of object-level entities, perspectives have been described by means of intensional principles (Simons 1987; Meirav 2003). Instrument semelfactive verbs such as *frustare* 'whip' in (35a), may be seen as denoting plural collections of events, and as such denote events with a potentially complex internal structure. In this case, we would like to suggest that the agent's commitment in performing a complex event, her being *sentient*, is precisely part of the intensional principle that enforces the perspective of taking a plurality of events of hitting as a unitary event *of a different type*. If we accept this

view on enforcing perspectives, it is possible to integrate in the general picture the observation we made about example (1b) at the very beginning, when we pointed out that the expression of the agent's intention, or of what may count as a reason in her/his eye, enhanced the processive reading of the predicate.

Despite their being perceptually accessible, the single sub-events cannot be considered fully referential, nor countable individuals. Such a multiplicity of sub-events does not count as a plurality. In this sense, instrument semelfactives are interpreted as *bona fide* activities, and nominalised semelfactives behave as well like intrinsically pluractional predicates of events, that is, as eventualities of a specific aspectual type.

4. Conclusions

In this paper, we discuss the semantics of semelfactive predicates by looking at *ata*-nominalisations built on instrument semelfactive verbs in Italian. First, we cast new light on the issue of what counts as a singular unit in the domain of semelfactives by comparing these units with the units in the denotation of *nomina vicis*. In either case, boundedness corresponds to a form quantisation independent of any prominent point in the event and/or in the theme. Second, we apply this line of exploration to the issue of the double reading – semel and processive – of semelfactives in Italian and their *ata*-nominalisations.

Ata-nominalisations are *nomina vicis* whose domain of denotation is restricted to singular events. Discretisation comes from aspectual boundedness, but the interpretation of the denoted events is also partly sensitive to the lexical properties of the base. *Ata*-nominalisations can be subdivided into three groups: those denoting an instance of activity, (*nuotata*), those with a semel reading, (*ombrellata*), and those that have mixed properties, (*bastonata*). In the third group we find instrument semelfactives, and we have provided double support for it being a separate group. First, the semel interpretation is not simply an entailment from the activity one, for the latter is missing in nominalisations derived from nouns. Second, the activity reading is not simply the pluralisation of the semel one, since it correlates with the presence of constraints on the instrument used to perform the event and strong entailments concerning the thematic realization of the external argument. We use the non-reducibility of this opposition to support the claim that semelfactives, in their processive reading, are not seen as the pluralisation of a single event, but have to be considered as *bona fide* activity predicates.

Acknowledgements

We gratefully acknowledge support by the project *DelimitEvent*, funded by *Fédération Typologie et Universaux du Langage* (CNRS FR 2559), and by Emerging Group *Dynamic Structuring in Language and Communication*, funded through the Institutional Strategy of the University of Cologne (ZUK 81/1). We thank the two anonymous referees for comments and criticisms that led to a reorganization of the paper.

References

Acquaviva, Paolo. 2005. "I significati delle nominalizzazioni in -ata e i loro correlati morfologici." In *La formazione delle parole*, ed. by Maria Grossmann and Anna Maria Thornton, 7–29. Roma: Bulzoni.

Aliquot-Suengas, Sophie and Yves Macchi. 2003. "Du même et de l'autre. Analyse contrastive de trois formes historiquement apparentées: les formes suffixales françaises *-ade/-ée* et la forme suffixale espagnole *-ada*." In *La linguistique hispanique dans tous ses états*. ed. by Christian Lagarde, 141–156. Perpignan: CRILAUP.

Arad, Maya. 2003. "Locality constraints on the interpretation of roots: The case of Hebrew denominal verbs." *Natural Language and Linguistic Theory* 21: 737–778. doi: 10.1023/A:1025533719905

Carlson, Gregory N. 1977. *Reference to Kinds in English*. Ph. D. thesis, University of Massachusetts at Amherst. Published by Garland, 1980.

Cusic, David. 1981. *Verbal plurality and aspect*. Ph. D. thesis, Stanford University.

Donazzan, Marta and Leticia L. Gritti. 2013. "Another look at telicity and homogeneity. Aspectual constraints on weak readings in two Romance languages." In *Time and TAME across languages*, ed. by Karina V. Molsing. Cambridge: Cambridge Scholar Publishing.

Dowty, David R. 1979. *Word meaning and Montague grammar*. Dordrecht: Reidel. doi: 10.1007/978-94-009-9473-7

Dowty, David R. 1991. "Thematic proto-roles and argument selection." *Language* 67: 547–619. doi: 10.1353/lan.1991.0021

Gaeta, Livio. 2000. "On the interaction between morphology and semantics: The Italian suffix -ata." *Acta Linguistica Hungarica* 47: 205–229. doi: 10.1023/A:1014066615802

Gatti, Tiziana and Lucia Togni. 1991. "A proposito delle interpretazioni dei derivati in *-ata* e in *s-*." *Arbeitspapiere des Fachbereichs Sprachwissenschaft*, University of Konstanz.

Hay, Jennifer, Chris Kennedy, and Beth Levin. 1999. Scalar structure underlies telicity in "degree achievements". In *Proceedings of Semantics and Linguistic Theory IX*, ed by Tanya Matthews and Devon Strolovitch, 127–144. Ithaca: Cornell University.

von Heusinger, Klaus. 2002. "The Interface of Lexical Semantics and Conceptual Structure. Deverbal and Denominal Nominalizations." *ZAS Papers in Linguistics* 27: 109–124.

Ippolito, Michela. 1999. "On the past participle morphology in Italian." *MIT Working Papers in Linguistics* 33: 111–137.

Kearns, K. 2007. Telic senses of deadjectival verbs. *Lingua* 117: 26–66. doi: 10.1016/j.lingua.2005.09.002

Kiparsky, Paul. 1982. "Word formation and the lexicon." *Proceedings of the Mid-America Linguistics Conference*, ed. by F. Ingemann, 3–29. Lawrence: Kansas.

Marantz, Alec. 2000. "Roots: The universality of root and pattern morphology." Handout of the talk delivered at Conference on Afro-Asiatic languages, University of Paris VII.

Meirav, Ariel. 2003. *Wholes, Sums, and Unities*. Dordrecht: Kluwer. doi: 10.1007/978-94-017-0209-6

Nesset, Tore. 2013. "The history of the Russian semelfactive: The development of a radial category." *Journal of Slavic Linguistics* 21: 123–169. doi: 10.1353/jsl.2013.0000

Rothstein, Susan. 2008. "Two puzzles for a theory of lexical aspect: the case of semelfactives and degree adverbials." In *Event Structures in Linguistic Form and Interpretation*, ed. by Johannes Dölling, Tatiana Heyde-Zybatow, and Martin Shäfer, 175–198. Berlin: Mouton De Gruyter. doi: 10.1515/9783110925449.175

Samek-Lodovici, Vieri. 2003. "The internal structure of arguments. Evidence from complex predicate formation in Italian." *Natural Language and Linguistic Theory* 21(4): 835–881. doi: 10.1023/A:1025588109815

Scalise, Sergio. 1984. *Generative Morphology*. Dordrecht: Foris.

Scher, Ana Paula. 2006. "Nominalizações em *-ada* em construções com o verbo leve *dar* em português brasileiro." *Letras de Hoje* 41: 29–48.

Simons, Peter. 1987. *Parts. A study in ontology*. Oxford: Clarendon Press.

Smith, Carlota. 1991. *The parameter of aspect*. Dordrecht: Kluwer. doi: 10.1007/978-94-015-7911-7

Tovena, Lucia M. 2010a. "Pluractional verbs that grammaticise number through the part-of relation." In *Romance Languages and Linguistic Theory 2008*, ed. by Reineke Bok-Bennema, Brigitte Kampers-Manhe, and Bart Hollebrandse, 233–248. Amsterdam: John Benjamins. doi: 10.1075/rllt.2.13tov

Tovena, Lucia M. 2010b. "Pluractionality and the unity of the event." In *The Amsterdam Colloquium 2009*, ed. by Maria Aloni and Karin Schulz, 465–473. Heidelberg: Springer.

Tovena, Lucia M. 2014. "Aspect and *-ata* event nouns in Italian." Handout of the talk delivered at ConSole XXII, Lisbon.

PART II

Morphosyntax

CHAPTER 5

Laísmo and "le-for-les"

To agree or not to agree

Adolfo Ausín and Francisco J. Fernández-Rubiera
Michigan State University / University of Central Florida

This paper discusses and analyzes the syntactic correlation between three different phenomena in Spanish: The presence of the accusative clitic, the presence/absence of number agreement in the dative clitic (a phenomenon called "le-for-les"), and the presence/absence of gender agreement in the dative clitic in *laísta* dialects. Assuming that accusative clitics manifest exclusively an agreement morpheme and that dative clitics can be decomposed into an applicative morpheme and an optional agreement morpheme, we argue for a unified account of these structures as follows: The presence of the accusative clitic and obligatory agreement in the dative are reflexes of the realization of an agreement morpheme in the structure. The analysis we propose also accounts uniformly for the unavailability of "la-for-las" or partial agreement in *laísta* dialects.

Keywords: Spanish clitics, applicative morpheme, agreement, "le-for-les", *laísmo*

1. Introduction

One of the features that distinguishes Spanish from other Romance languages such as French and Italian is the possibility of having clitic doubling constructions where an argument appears both as a clitic and as a full DP as illustrated abstractly in (1) for dative arguments.

(1) Dative clitic doubling in Spanish
LE(S) Verb… Dative DP
(Dat CL)

The availability of clitic doubling is subject to a number of factors: type of argument (dative or accusative), type of full DP (pronominal or non-pronominal), type of reference (definite or indefinite), and dialect (some dialects being more permissive than others). All of these factors have received a lot of attention in the literature

DOI 10.1075/rllt.11.05aus

and we will come back to some of them. One aspect that has not received much attention is the agreement between the clitic and its double. It is normally assumed without any debate that the clitic agrees with the doubling DP, as illustrated in (2) where the plural dative DP *a los estudiantes* agrees with the plural dative clitic *les*.

(2) *Les dimos un libro a los estudiantes* (Normative Spanish)
LES gave1PL a book to the students
'We gave a book to the students.'

However, the dative clitic and the dative DP do not always agree. For instance, a dative clitic (in bold) may or may not show number agreement with the DP that it doubles (underlined) in most varieties of Spanish, as the examples in (3) below illustrate:

(3) Dative number agreement
a. *Dale cuerda a tus oportunidades, la Lotería reparte más*
give-LE rope to your opportunities the Lottery delivers more
de 5 millones
than 5 million
'Wind up your opportunities, the lottery gives more than 5 million.'
(Huerta Flores 2005)

b. *A los muchachos* ***les/*le*** *compré una alarma de coches*
to the guys LES/LE bought1SG an alarm of cars
contra robos
against thefts
'Those guys, I bought them a burglar alarm for the car.'
(Huerta Flores 2005)

The use of the non-agreeing *le* in (3a) instead of the agreeing *les* has been labeled "le-for-les" in the descriptive grammatical tradition, a term that we will adopt in this paper.

Furthermore, we find a similar lack of agreement between the clitic and its double in *laísta* dialects. The defining feature of *laísta* dialects is that dative clitics show a masculine/feminine alternation. However, as discussed by Romero (2012) and shown in (4), gender agreement in clitic doubling constructions between the clitic and its double is dispreferred in these dialects.

(4) Dative gender agreement in *laísta* dialects
a. *Cuando vi a Marta,* ***la*** *di un libro*
when saw1SG to Marta, LA gave1SG a book
'When I saw Marta, I gave her a book.'

b. ***le/??la*** *di* *un* *libro* *a* *Marta*
LE/LA gave1SG a book to Marta
'I gave Marta a book.'

Since in (4b) there is a preference for the non-gender agreeing *le* over the gender agreeing *la*, we would like to label this phenomenon "le-for-la".

These two phenomena (i.e. "le-for-les" or lack of number agreement and "le-for-la" or lack of gender agreement) have not been previously related in the literature, and our main goal in this paper is to provide an analysis that explains these structures uniformly. We will do so by relating these two phenomena to the presence of accusative clitics. To do so, we will argue that the dative clitic is a composite of two different morphemes: an applicative morpheme and optionally an agreement one. We will claim that the presence or absence of number agreement in (3) and gender agreement in (4) can be accounted for under the same terms: agreement (number for most varieties of Spanish and number and gender for *laísta* varieties) obtains when the applicative morpheme is combined with the agreement morpheme; lack of agreement results when only the applicative morpheme surfaces. We will relate the presence or absence of agreement in dative clitics with the possibility of having an accusative clitic. We will show that the contexts that favor agreement in dative clitics are the same contexts where accusative clitics are required. The summary of our proposal is shown below:

(5) Possible realizations of dative clitics
a. Applicative + Agreement: agreeing *les* and *la*: (3b) and (4a).
b. Applicative only: non-agreeing *le* ("le-for-les" and "le-for-la"): (3a) and (4b)

This paper is organized as follows. Section 2 discusses the absence of number agreement with the dative (i.e. "le-for-les") and correlates this phenomenon with the presence of an accusative clitic, showing how our proposal accounts uniformly for the presence/absence of agreement with the dative and the presence/absence of the clitic with the accusative. Section 3 presents and discusses *laísta* Spanish dialects in light of Romero's (2012) analysis. In Section 4 we will show how our proposal can account for the lack of gender agreement in dative clitics discussed by Romero. In the last section we lay out our conclusions and some possible future research ideas.

2. "Le-for-les" and the presence of accusative clitics: A common core

2.1 Background on "le-for-les" constructions

The use of singular *le* instead of the expected plural *les* has been acknowledged, documented, and discussed by Cuervo (1907, 1955), Casares (1918: 107–120), Sturgis (1927), Rinni (1988), DeMello (1992), Fernández-Soriano (1999), Huerta Flores (2005), RAE/AALE (2009: 2664), among others.[1] To show how widespread the phenomenon was, Cuervo noted the example in (6) from the Spanish Royal Academy (RAE). The intended purpose of this example was to illustrate cacophony. The unintended effect was to illustrate the use of "le-for-les" even among the most educated speakers.

(6) *Dale las lilas a las niñas*
give-LE the lilacs to the girls
'Give the lilacs to the girls.' (RAE 1885: 287)

As observed by many, it is not the case that "le-for-les" is acceptable in any syntactic context. "Le-for-les" has been shown to occur mostly with postverbal dative phrases, as in (7), but barred when the dative phrase is either dislocated, as in (8a–c), or null as in (8d). We will come back to these contrasts later in the paper.

(7) "Le-for-les" contexts
a. ***Le*** *dice adiós á las garzas que pasan*
LE_{SG} say3SG goodbye to the herons that fly-by
'S/he says goodbye to the herons that fly-by.'
b. *Yo no* ***le*** *tengo miedo á las balas*
I not LE_{SG} have1SG fear to the bullets
'I am not scared of bullets.'
c. ***Le*** *dice a todos que vengan*
LE_{SG} say3SG to all that come3PL
'S/he tells everyone to come.' (Cuervo 1907: §309)
d. *Todo se acabaría si* ***le*** *pegaran cuatro tiros a unos cuantos granujas*
All rflCL would-end3SG if LE_{SG} shoot3PL four shots to some few crooks
'It would all be over if they shot dead a few crooks.' (Casares 1918: 108)

1. See Boeckx and Jeong (2004) for a generative approach to this phenomenon along completely different lines from ours.

(8) Ungrammatical "le-for-les" contexts

a. *á las garzas {*le/les} dice adiós*
to the herons LE/LES say3SG goodbye
'The herons, s/he says goodbye to them.'

b. *á las balas yo no {*le/les} tengo miedo*
to the bullets I not LE/LES have1SG fear
'Bullets, I am not scared of them.'

c. *a todos {*le/les} dice que vengan*
to all LE/LES say3SG that come3PL
'Everyone, s/he tells them to come.' (Cuervo 1907)

d. *Aquí hay dos caballeros que desean ver al señorito.*
Here are two gentlemen that wish3PL seeINF to-the master
*¿Qué {*le/les} digo?*
What LE/LES say1SG
'Here are two gentlemen who wish to see the master. What should I tell them?' (Casares 1918)

Cuervo describes the lack of agreement between the dative clitic and its double in (7) as a "brilliant mistake" (*error genial*) but without making explicit why he thinks it is brilliant. Casares (1918: 108–109) also considers the use of "le-for-les" a mistake, but he shows forgiveness and understanding. Thus, regarding (7d), he notes that it "sounds more spontaneous and natural with the defective *le* than with the *les* that the grammar requires". Casares not only shows sympathy for the phenomenon under consideration but he also sketches an explanation. For him, the difference between (7d) and (8d) relates to the clitic's function. In structures such as those in (7d), Casares claims that the clitic is not functioning as a pronoun that refers to an antecedent, but rather as a kind of "expletive or adverbial particle". In Casares' words, the clitic "becomes a little arrow pointing forward to indicate the presence of an indirect object".[2]

We essentially agree with Casares's analysis and would like to relate his "little arrow" view to the applicative morpheme analysis of dative clitic constructions developed by M. Cristina Cuervo (2003, 2010). We would like to argue that the non-agreeing *le* is the materialization of the applicative morpheme, whereas the agreeing *les* is the materialization of both an applicative morpheme and an agreement morpheme. What needs to be discussed are the contexts where each option is realized. As a first step, we will note the parallelism between the contexts that favor the agreeing *les* and those that favor the presence of an accusative clitic in the next section.

2. Translations are our own.

2.2 "Le-for-les" and the presence of accusative clitics: Evidence for a uniform treatment

As we saw earlier in (8), there are some syntactic contexts where "le-for-les" is not possible. The purpose of this section is to show the connection between doubling structures with the dative and the presence of accusative clitics. In short, the syntactic environments that exhibit obligatory number agreement with the dative correlate with those where the accusative clitic must be present. Our ultimate goal is to motivate a uniform treatment of these two previously unrelated phenomena.

Roughly speaking, accusative clitics are required in the very same syntactic environments where "le-for-les" is not possible. One such context is illustrated in (9). As it is shown, a left-dislocated object (indirect or direct) forces the presence of obligatory agreement with the dative and the presence of the clitic with the accusative, as in (9a) and (9b) respectively.

(9) a. *A las balas yo no {*le/les} tengo miedo*
to the bullets I not LE/LES have1SG fear
'Bullets, I am not scared of them.'

b. *A Marta Juan *(la) vio*
to Marta Juan her$_{CL}$ saw3SG
'Marta, Juan saw her.'

A second context that illustrates the parallelism between these two phenomena is that of covert objects (i.e. *pro*). In these instances, both obligatory agreement with the dative and an obligatory clitic with the accusative arise, as in the examples below.

(10) a. *Aquí hay dos caballeros que desean ver al señorito.*
here are two gentlemen that wish3PL see$_{INF}$ to-the master
*¿Qué {les/*le} digo pro$_{[PL]}$?*
What LES/LE say1SG *pro*
'Here are two gentlemen who wish to see the master. What should I tell them?'

b. *Juan *(la) vio pro*$_{[FEM, SING]}$
Juan her$_{CL}$ saw3SG *pro*
'Juan saw her.'

A third context that illustrates this parallelism between obligatory number agreement in the dative and an obligatory clitic with the accusative is that of overt pronouns. Similar to those cases discussed above, the dative must show agreement and the accusative clitic must be present. Consider (11):

(11) a. *Juan {les/*le} entregó un libro a ellos*[3]
Juan LES/LE gave3SG a book to them
'Juan gave a book to them.'

b. *Juan *(la) vio a ella*
Juan her$_{CL}$ saw3SG to her
'Juan saw her.'

The descriptive generalization that seems to emerge from these facts is that the contexts that do not allow "le-for-les" (i.e. lack of number agreement with the dative) are the same as the contexts where an accusative clitic is required. The next subsections provide an account for this descriptive generalization. In 2.3 we will propose that dative clitics should be decomposed into an applicative morpheme and an agreement morpheme. In 2.4 we will show that this proposal can account for the uniform behavior illustrated above regarding obligatory number agreement with the dative and the obligatory presence of an accusative clitic.

2.3 Main assumptions and proposal

First, following Demonte (1995) and Cuervo (2003), we assume the dative clitic *le* plays a crucial role in the argument structure of the verb. For Demonte, the presence of the dative clitic triggers an alternation similar to that found in double object constructions in English; that is, the contrast between the examples in (12) in Spanish would be similar to the contrast found in those in (13).

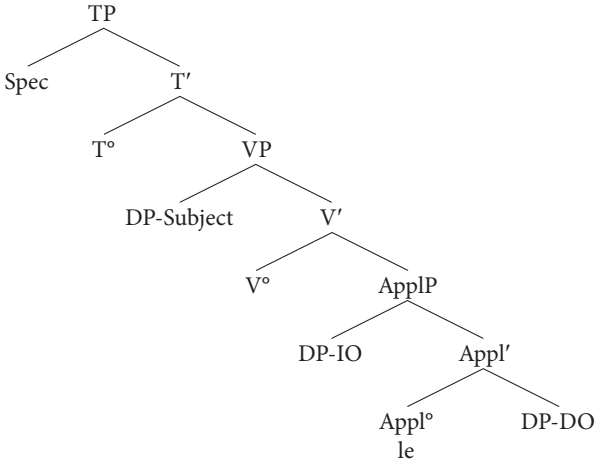
(12) a. *Entregué las llaves al conserje*
b. *Le entregué las llaves al conserje*
(CL) gave1SG the keys to-the janitor
'I gave the keys to the janitor.'

(13) a. *John gave a book to Mary*
b. *John gave Mary a book*

M. Cristina Cuervo (2003, 2010) takes up Demonte's ideas and develops an applicative analysis of dative clitic constructions. She proposes that the dative clitic in Spanish is the realization of an applicative morpheme that introduces an extra, applied argument. The underlying structure that M. C. Cuervo proposes for dative clitic constructions is shown in (14).

3. The grammaticality contrast in (11a) is probably not as strong as in (9a) and (10a). We will abstract away from this difference and leave it open for future research.

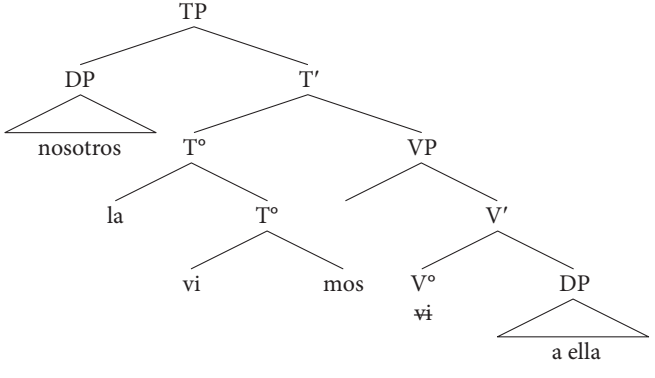
(14) M. C. Cuervo (2003, 2010): Applicative analysis of dative clitic constructions



Second, following Suñer's (1988) seminal work on clitics, we assume that accusative clitics instantiate an agreement morphological marker similar to subject agreement morphemes. Thus, a sentence as in (15) receives the analysis in (16).[4]

(15) *Nosotros la vimos a ella*
we her$_{CL}$ saw1PL to her
'We saw her.'

(16)



4. In line with Suñer's analysis in (16), we assume that the accusative clitic manifests object-agreement realized in T°, which in turn could be analyzed as an independent head in the structure. However, we do not explicitly indicate what the status of the clitic is and leave this issue here for future research. See also Sportiche (1996) for a more complex development of clitics as agreement markers.

Finally, combining Cuervo's analysis of dative clitics as applicative morphemes and Suñer's proposal of accusative clitics as agreement markers, we argue for a decomposition of dative clitics into an obligatory applicative morpheme and an optional agreement morpheme, as illustrated in (17).[5]

(17) Dative clitics = Applicative morpheme (+ Agreement morpheme)

In other words, we are proposing that accusative and dative clitics are fundamentally different. Whereas accusative clitics are agreements markers, dative clitics are applicative morphemes that can also contain an agreement morpheme.[6]

In what follows, we will show how the previous assumptions can account uniformly for the presence of an accusative clitic and for "le-for-les".

2.4 Accounting for "le-for-les" and the presence of accusative clitics

Let's consider first the examples in (18). In these examples the double DP is an overt pronoun. As it is shown, overt pronouns trigger both obligatory agreement with the dative and the obligatory presence of an accusative clitic:

(18) a. ***Les/*Le*** *dimos* *a* *ellos* *un* *libro*[6]
LES/LE gave1PL to them a book
'We gave them a book.'

b. *Nosotros* ****(la)*** *vimos* *a* *ella*
we her$_{CL}$ saw1PL to her
'We saw her.'

Following Suñer's assumption that accusative clitics should be analyzed as agreement morphemes, the syntactic representation of (18b) would be (16), repeated here as (19) with minor variations. Since the direct object DP is an overt pronoun, the agreement marker shows up and is realized as a clitic.

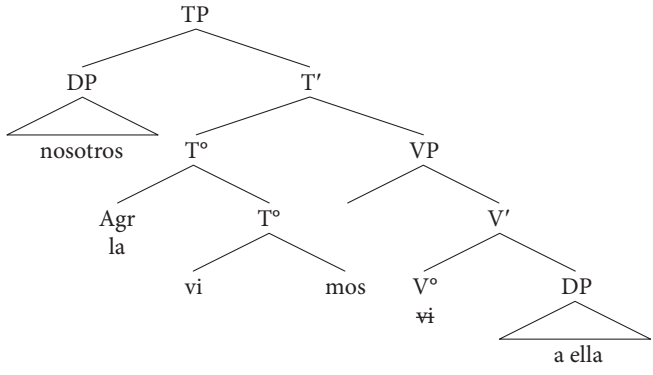
---

5. Alternatively, we could assume that dative clitics can be lexically decomposed into an agreement morpheme and a deictic morpheme as proposed in Martín (2012).

6. The idea that dative and accusative clitics are different has many antecedents in the literature. For instance, Uriagereka (1988), Bleam (1999), Ormazabal and Romero (2013), Huerta Flores (2005) among others assume that dative clitics are agreement morphemes, but accusative clitics are something else (a determiner or a pronoun, depending on the author).

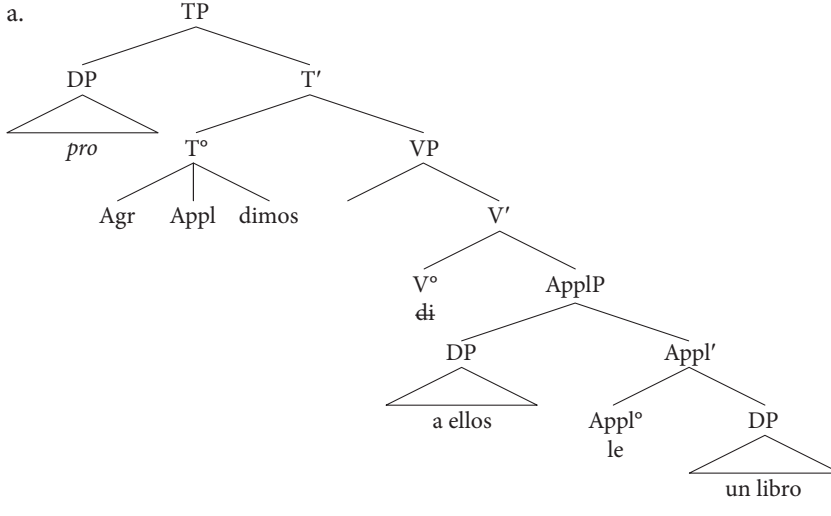
7. Since the relative order of the DO and IO does not seem to be relevant (Huerta Flores 2005: fn11), we will be using either without any discussion, occasionally choosing the one that is more convenient for expository purposes.

(19)



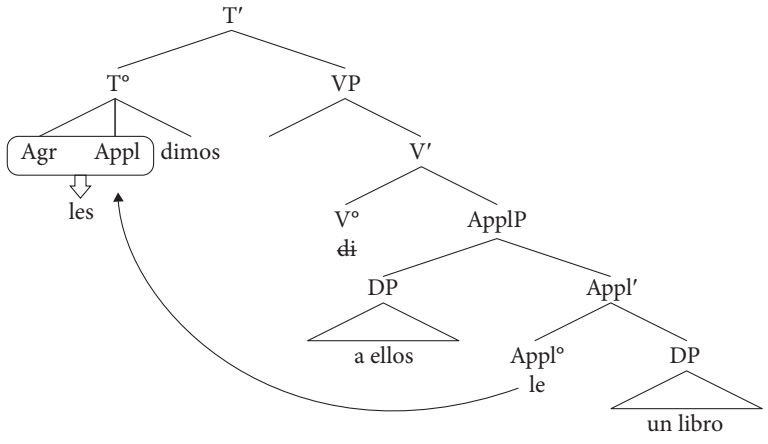
Let's turn our attention to (18a). In this example we have a dative clitic doubling a covert pronoun. In line with Cuervo's analysis of dative clitics in (14), we assume that the dative clitic in (18a) is initially merged as an applicative head, as in (20a). Furthermore, since in this case the doubling DP is a pronoun, there needs to be present an agreement morpheme which combined with the applicative morpheme results in the agreeing *les*, as shown in (20b).[8]

(20) a.



8. For expository purposes, we are placing agreement under T. In a more detailed representation it should head its own projection as in Suñer (1988) and Sportiche (1996), among others.

b.



In other words, if we assume that strong pronouns in canonical position (*a ellos* in (18a) and *a ella* in (18b)) trigger the presence of an agreement morpheme, we can provide a uniform account of the clitic patterns in the said examples. There is number agreement with the dative in (18a), and there is an accusative clitic in (18b).

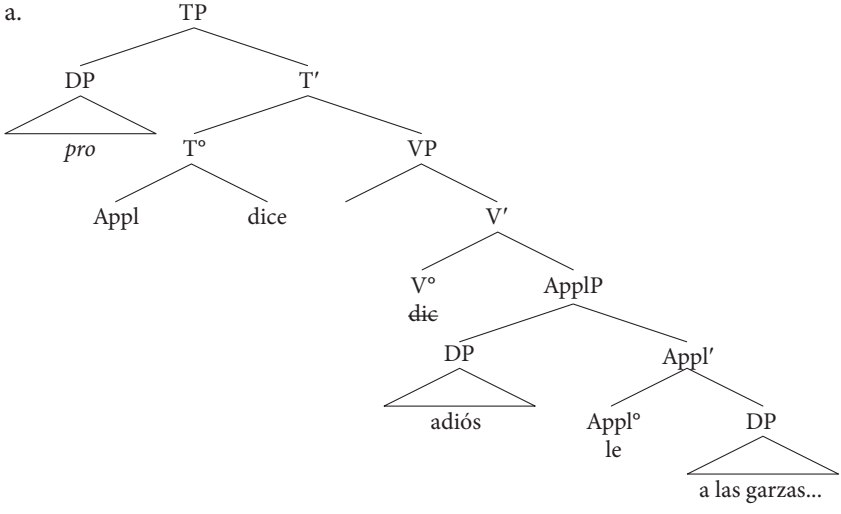
If the DP in canonical position is not a pronoun as in (18) but a non-pronominal DP as in (21), then the clitic patterns change: the non-agreeing *le* is acceptable as illustrated in (21a), and there is no accusative clitic doubling as illustrated in (21b). We would like to account for both cases in the same way: non-pronominal DPs do not trigger the presence of agreement. In the accusative case, this lack of agreement results in the absence of an accusative clitic in (21b). In the dative case, the lack of agreement results in an instance of "le-for-les" in (21a).[9]

(21) a. ***Le*** *dice* *adiós* *a* *las* *garzas* *que* *pasan*
LE say3SG goodbye to the herons that fly-by
'S/he says goodbye to the herons that fly-by.'

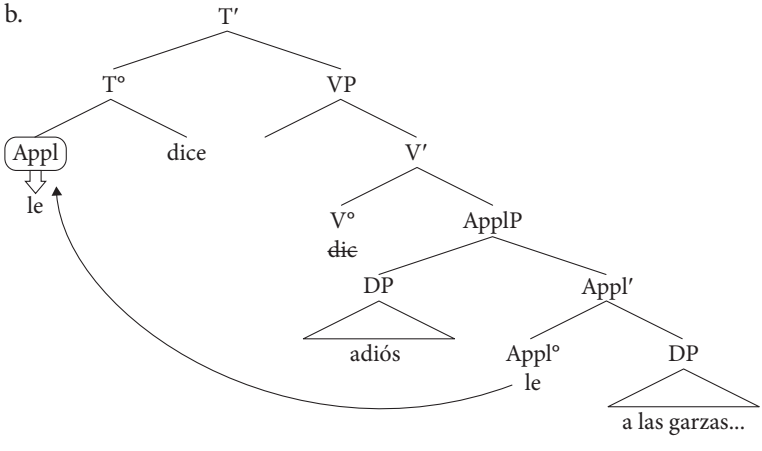
9. We are assuming that in examples like (21a) the grammatical option is the non-agreeing *le*. It is true that the agreeing *les* is not ruled out in colloquial Spanish. We believe that there are several factors involved. One is normative pressure: it could be that the presence of *les* in normative Spanish might be favoring the presence of agreeing *les* in colloquial Spanish too, in what could be considered a case of hypercorrection in these speakers' internal grammar. Another possible factor is dialectal variation: in the same way that we find dialectal variation in accusative clitic doubling (Suñer 1988), it could be possible that there is also dialectal variation in the use of "le-for-les". Finally, it could also be possible that there are some discourse and informational factors that favor one option or the other (Huerta Flores 2005). In this paper we will abstract away from these possible factors. What is crucial for our proposal is the fact that the disagreeing *le* is also attested, which together with its systematicity and pervasive nature in most dialects of Spanish call for an analysis of this instance of a non-agreeing clitic with the dative.

b. *(*Las) mira las garzas que pasan*
Them$_{CL}$ look3SG the herons that fly-by
'S/he looks at the herons that fly-by.'

(22) a.



b.



The agreeing form *les* and the accusative clitic (i.e. in our terms, Agreement) must obligatorily show up in cases as those in (23). If the dislocated element *a las garzas* – "to the herons" – is base-generated in preverbal position, with *pro* occupying the complement slot,[10] the obligatory presence of the agreeing *les* and *la* can be accounted for under the same terms as in (18): pronouns (null or overt) force the

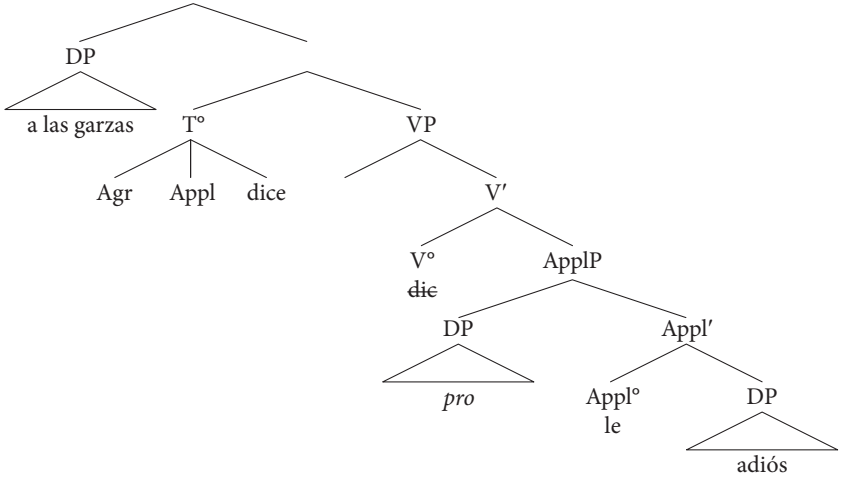
10. We are assuming a base generation approach to Clitic Left Dislocation. For discussion, alternatives, and references, see Alexiadou (2006), among many others.

presence of the agreeing head, as a result of which *les*, and not *le*, shows up. The analysis for (23a) is shown in (24).

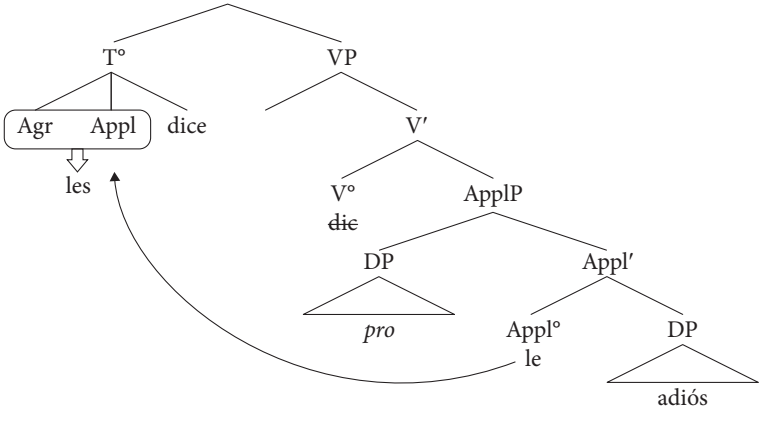
(23) a. *A las garzas les dice adiós*
to the herons LES say3SG goodbye
'The herons, s/he says goodbye to them.'

b. *A Marta Juan *(la) vio*
to Marta Juan her$_{CL}$ saw3SG
'Marta, Juan saw her.'

(24) a.



b.



Under our proposal, the presence of a left dislocated element and *pro* in argument position in the examples in (23) triggers the presence of the agreement morpheme. This agreement is responsible for the appearance of the accusative clitic in (23b) and the agreeing *les* in (23a). As shown in (24), the applicative morpheme moves to the inflectional domain where it merges with the agreement morpheme and both are realized as the agreeing *les*.

One could argue that the analysis depicted in (22) and (24) is somehow stipulative: if there is an agreeing dative clitic, we assume the presence of an agreement morpheme as in (22) but if a dative clitic does not agree, we propose that there is no agreement morpheme. However, this potential criticism does not take into account that the same mechanism is used to explain the presence of accusative clitics. In other words, the criticism would only be valid if there were no independent empirical motivation for our proposal, which is not the case. Furthermore, in the next section we will show that the very same proposal can also account for the lack of gender agreement in *laísta* dialects.

In the next section, we will present the phenomenon called *laísmo* and review Romero's account. We will ultimately conclude that (i) Romero's analysis falls short at accounting for the data we will discuss, and (ii) our decomposition analysis of dative clitics is amenable to be extended to account uniformly for these two previously unrelated phenomena, namely "le-for-les" and *laísmo*.

3. Lack of agreement in *laísta* dative clitics

In this section, we will discuss a different case where dative clitics do not seem to agree with their doubles. In this case, we will be discussing lack of gender agreement in *laísta* dialects. As far as we know, these facts have been noted first by Romero (2012). In the following, we will introduce *laísmo*, and we will discuss Romero's analysis. In the next section we will show how the proposal that we developed in the previous section for "le-for-les" can also account for the lack of gender agreement that *laísta* dialects of Spanish may manifest.

The Spanish clitic system is subject to a great deal of variation. In the standard system, there is a masculine/feminine distinction in the accusative but not in the dative as illustrated in (25). In Castilian Spanish, shown in (26), it is normal to find

a further distinction in the masculine accusative: *lo* is used for inanimates and *le* (typically, the dative clitic)[11] for animates. This is what is called *leísmo*.[12]

(25)	**Standard Spanish**	Accusative	Dative
	Singular	Masc.: *lo*	*Le*
		Fem: *la*	
	Plural	Masc.: *los*	*Les*
		Fem: *las*	

(26)	**Castilian Spanish (*Leísta*)**	Accusative	Dative
	Singular	**Masc. Animate: *le***	*Le*
		Masc. Inanimate: *lo*	
		Fem: *la*	
	Plural	**Masc. Animate: *les***	*Les*
		Masc. Inanimate: *los*	
		Fem: *las*	

A third dialect, the *laísta* dialect, is illustrated in (27). It shows a gender difference in the dative. As can be seen in (25) and (26), dative clitics are not marked for gender in most dialects. However, *laísta* dialects (which tend to be *leísta* too) show a gender difference in the dative clitic: *le* is used for masculine and *la* (typically, the accusative clitic) for feminine. The contrast between the clitic systems in (25) and (27) is illustrated in the examples in (28) and (29) respectively.

(27)	*Leísta* and *laísta* Spanish	Accusative	Dative
	Singular	**Masc. Animate: *le***	Masc.: *Le*
		Masc. Inanimate: *lo*	**Fem: *La***
		Fem: *la*	
	Plural	**Masc. Animate: *les***	Masc.: *Les*
		Masc. Inanimate: *los*	**Fem: *Las***
		Fem: *las*	

11. Contrary to Bleam (1999) and others, we assume that dative *le* and accusative *le* in leísta dialects are different. A piece of evidence in favor of this is that while *le-for-les* is possible with dative *le*, it is not with accusative *le*. Many *leísta* speakers accept accusative clitic doubling as in (i). However, as far as we know, *le-for-les* in contexts like this is not possible, as illustrated in (ii):

(i) Les vi a los niños.
LES.Acc saw.I to the children

(ii) *Le vi a los niños
LE.Acc saw.I to the children
'I saw the children'

12. For more information and references on *leísmo* and *laísmo*, see Fernández-Ordóñez (1994).

(28) *Cuando vi a Marta,* ***le*** *di un libro* (Standard Spanish)

(29) *Cuando vi a Marta,* ***la*** *di un libro* (*Laísta* Spanish)
when saw1SG to Marta, CL gave1SG a book
'When I saw Marta, I gave her a book.'

Most descriptions of *laísta* dialects assume that dative clitics are always realized as *la* independently of any configurational factors. For instance, the *Nueva Gramática* (RAE-ASALE 2009, 1:6.10a) defines *laísmo* as "the use of pronoun *la* as dative with feminine nouns." Two relevant examples appear in (30).

(30) a. ***La*** *dije la verdad*
LA said the truth
'I told her the truth.'

b. *No te puedes poner esta camisa porque tengo que pegar****la*** *un par de botones*
not CL can put that shirt$_{FEM}$ because have to attach.CL$_{FEM}$ a couple of buttons
'You cannot wear that shirt because I need to sew a couple of buttons on it.'

This simplistic view of *laísmo* has recently been challenged by Romero (2012). Romero points out that in (at least some) *laísta* dialects, the agreeing form *la* in the dative is disfavored when there is clitic doubling as illustrated in (31).

(31) *Le/ ??La envié un regalo a la niña*
LE LA sent a gift to the girl
'I sent the girl a gift.'

Romero explains the impossibility of having *laísmo* in (31) by assuming that the dative *la* is actually an accusative clitic. Since accusative clitics typically do not allow clitic doubling, the ungrammaticality of the clitic *la* in (32) is extended, according to Romero, to that in (31).

(32) **La vimos a la niña*
LA saw to the girl
'We saw the girl.'

Romero bases his claim that dative *la* is actually an accusative on the belief that *la* as a dative is not possible on those structures that do not license accusative, as in (33). Romero claims that since in these structures accusative case is not licensed, dative *la* (an accusative clitic for him) does not arise in these structures.

(33) a. *El regalo le/*la fue enviado*
the present LE/LA was sent
'The present was sent to her.'

b. *La carta le/*la llegó tarde*
the letter LE/LA arrived late
'The letter arrived late for her.'

c. *Le/*la parecía estúpido*
LE/LA seemed stupid
'It seemed stupid to her.'

Romero finds additional evidence for his proposal (that is, *la* in dative constructions is an accusative) on some alleged animacy constraints on both accusative case assignment and *laísmo*. He claims that both are subject to an animacy constraint. As for accusative case, Romero assumes that it is only assigned to animate direct objects like the one in (34a) but not to inanimate objects like the one in (34b). As for *laísmo*, he uses the contrasts in (35) to support the claim that *laísmo* is possible only with animates.[13]

(34) a. *Busco a Marta*
search1SG to Marta
'I am looking for Marta.'

b. *Busco la respuesta*
search1SG the answer
'I am looking for the answer.'

(35) a. *Le/*La puse los tornillos (a la mesa)*
LE/LA put the screws to the table
'I attached the screws to the table.'

b. *Le/*La añadí azúcar (a la leche)*
LE/LA added sugar to the milk
'I added sugar to the milk.'

We are skeptical of Romero's view that dative *la* is actually an accusative for two reasons. First, as the *Nueva Gramática* explicitly states, *laísmo* is possible with intransitive verbs; that is, it is possible with verbs that do not assign accusative case.[14] Some of the examples that they discuss appear in (36).

---

13. See Romero (2012: §3.2) for more details.

14. It seems to us that the verbs in these examples are clear cases where there is no accusative case available in the structure. The examples in (36a) (*salir* 'to turn out') and in (36b) (*hincharse*

(36) Examples of *laísmo* with intransitive verbs from the *Nueva Gramática*
  a. *Hoy* ***la*** *sale todo a Carmen Corcelles*
     'Carmen Corcelles makes everything right today.'
  b. *Los pies se* ***la*** *hinchan*
     'Her feet swell up.'
  c. ***La*** *gusta el rollo*
     'She loves flirting.'
  d. *¿Qué* ***la*** *duele?*
     'What hurts her?'

Second, it is not clear to us that *laísmo* is restricted to animates, as Romero claims. The *Nueva Gramática* uses the examples in (37) to show that it is possible to find *laísmo* with inanimates.[15]

(37) Examples of *laísmo de cosa* from the *Nueva Gramática*
  a. *No te puedes poner esta camisa porque tengo que pegar****la***
     Not CL can put that shirt$_{FEM}$ because have to attach.LA
     un par de botones
     a couple of buttons
     'You cannot wear that shirt because I need to sew a couple of buttons on it.'
  b. *Si a la tortilla se* ***la*** *añade demasiada cebolla*
     if to the omelet SE LA adds too much onion
     'If they add too much onion to the omelet…'

In conclusion, out of the three features that Romero assigns to *laísmo* (namely, lack of doubling, restricted to animates, and restricted to "accusative contexts") we are skeptical about the last two. As for the first one, we will devote the rest of the paper to show that the use of *le* for *la* in *laísta* dialects of Spanish (hereafter, "le-for-la") is also amenable to an analysis along the lines of what we proposed to account for "le-for-les" in Section 2 above.

'to swell') are unaccusative and those in (36c) (*gustar* 'to like') and (36d) (*doler* 'to hurt') are non-agentive psych verbs.

15. The *Nueva Gramática* does say that *laísmo de cosa* (i.e. *laísmo* used to refer to inanimate objects) is less common than *laísmo de persona* (i.e. *laísmo* used to refer to animate objects). However, this could just be due to the fact that in general, it is much more common to find animate datives than inanimates ones. It could also be the case that because there are fewer cases of inanimate datives, they are more susceptible to normative pressure.

4. A uniform account of "le-for-les" and "le-for-la"

The contrast that we are trying to explain is shown in (31), repeated here for convenience. What must be explained is why (gender) agreement is not manifested in the clitic in this particular instance.

(38) *Le/ ??La envié un regalo a la niña*
LE LA sent a gift to the girl
'I sent the girl a gift.'

Remember that Romero explains the contrast in (38) by proposing that the dative *la* is actually accusative and as such, it cannot participate in clitic doubling constructions. We, on the other hand, would like to relate the contrast in (38) to the "le-for-les" phenomenon. In order to do that, we will first show the parallelism between the two constructions, and then we will provide an explanation for (38) based on our account of "le-for-les."

4.1 Parallelism between "le-for-les" and "le-for-la"

There is a strong parallelism between successful instances of *laísmo* and obligatory agreement with datives. *Laísmo* tends to be available only in those contexts where dative agreement is required. In the same way, there is a strong parallelism between "le-for-les" and "le-for-la". That is, the same contexts that favor "le-for-les" also favor "le-for-la."

Let's start with clitic doubling with full DPs. As discussed in Romero (2012), *laísmo* is not allowed in clitic doubling constructions with lexical DPs. In other words, clitic doubling constructions favor "le-for-la." In the same context, agreement is disfavored in most registers of Spanish, meaning "le-for-les" is preferred. This is illustrated in (39).

(39) Clitic doubling with lexical (non pronominal) DP
a. *Le/??La envié un regalo a la niña*
'We sent the girl a gift.'
b. *Le dimos las lilas a las niñas*
'We gave the girls the lilacs.'

Now let's consider the contexts where both *laísmo* and agreement are favored in the relevant dialects. That is, these are the contexts where both "le-for-la" and "le-for-les" are not acceptable. Let's start with left dislocation structures. As shown in (40), in left dislocation structures, the agreeing options are preferred: both "le-for-la" and "le-for-les" are not acceptable.

(40) Left dislocation
  a. *A Marta la/??le hemos comprado un libro*
     'Marta, we bought her a book.'
  b. *A los niños les/*le hemos comprado un libro*
     'The kids, we bought them a book.'

The examples in (41) are cases where there is a null object (*pro*) in object position. As in the previous examples, the agreement option is preferred: "le-for-la" and "le-for-les" are not acceptable.

(41) Null object (pro) in object position
  a. *Cuando hables con Marta, dala/??le recuerdos de mi parte*
     'When you talk to Marta, give her my regards.'
  b. *Cuando hables con tus padres, dales/*le recuerdos de mi parte*
     'When you talk to your parents, give them my regards.'

We find the same situation in examples with clitic doubling with a pronominal double as in (42). Again, the agreement option is preferred: "le-for-la" and "le-for-les" are not acceptable.[16]

(42) Clitic doubling with a pronominal double
  a. *La/??Le hemos comprado un libro a ella*
     'We bought her a book.'
  b. *Les/??Le hemos comprado un libro a ellos*
     'We bought them a book.'

The parallelism between these two constructions appears summarized in the following table:

(43) Summary

Context	***Laísmo* available / "le-for-la" unacceptable**	**Dative agreement required / "le-for-les" unacceptable**
Left dislocation	A Marta la hemos comprado un libro	A los niños les hemos comprado un libro
Null object	La hemos comprado un libro	Les hemos comprado un libro
Pronominal Double	La hemos comprado un libro a ella	Les hemos comprado un libro a ellos

16. In the a. examples in (40)–(42), the option with *le* is not completely ruled out even for *laísta* speakers such as one of the authors, a predominantly *laísta* speaker from Madrid. We suspect that this is due to pressure from the more prestigious, non-*laísta* dialects. We leave this issue open for future research.

Context	"le-for-la" preferred	"le-for-les" preferred
Lexical DP Double	Le/ ??La envié un regalo a la niña	Le/??Les compramos un libro a los niños

In order to account for this parallelism we would like to suggest that the availability of "le-for-la" falls under the same generalization as "le-for-les." In the dialects under consideration, gender is part of the agreement features that can be expressed with applicative morphemes. Thus, in *laísta* dialects an applicative morpheme plus agreement results in *la(s)*. In other words, the *laísta* pattern just reviewed can be derived from the proposal in (17) repeated below. The difference between Standard Spanish and *laísta* dialects of Spanish would be that in Standard Spanish, dative agreement involves only number agreement, but in *laísta* Spanish, dative agreement involves both number and gender features.

(17) Dative clitics = Applicative morpheme (+ Agreement morpheme)

If we apply this proposal to the examples in (44) and (45), we would say that in (44) the clitic *la* is the materialization of the applicative morpheme and the gender and number features in *laísta* Spanish, whereas in (45) the clitic *le* would be the materialization of just the applicative morpheme in both *laísta* and non-*laísta* Standard Spanish.

(44) *Cuando vi a Marta, **la** di un libro* (*Laísta* Spanish)
when saw to Marta, CL gave a book
'When I saw Marta, I gave her a book.'

(45) ***Le** envié un regalo a la niña* (Standard AND *laísta* Spanish)
CL sent a gift to the girl
'We sent the girl a gift.'

Differently put, what we are proposing is that "le-for-les" and what we have called "le-for-la" are essentially the same phenomenon. In both cases, the non-agreeing option *le* is the result of the applicative morpheme being materialized on its own, without an agreement morpheme. This proposal straightforwardly explains the parallelism in (43): "le-for-les" and "le-for-la" appear in the same syntactic contexts because those are the contexts where the agreement morpheme is not present in the syntactic structure.

4.2 A further prediction: Lack of "la-for-las"

Under the current proposal, we have a different take to account for the ungrammaticality of *la* instead of *las* in dative constructions in *laísta* dialects. Fernández Soriano (1999) observed that examples like (46) and (47) are not attested in *laísta* dialects.

(46) **La tiene miedo a las balas*
LA has fear to the bullets$_{\text{FEM}}$
'S/he is afraid of bullets.'

(47) **Nunca la agrada a las jefas la disconformidad de las empleadas*
never LA pleases to the bosses$_{\text{FEM}}$ the disconformity of the employees
'The employees' disconformity never pleases the bosses.'

According to Fernández Soriano, the descriptive generalization that captures examples like (46) and (47) is that the equivalent of "le-for-les" is not possible in *laísta* constructions; that is to say, *laísta* dialects of Spanish do not exhibit what we may call "la-for-las." We agree with the evidence presented by Fernández Soriano, and we also believe that these facts follow from our proposal under the reasonable assumption that gender and number features are packed in the same agreement morpheme. If the syntactic structure requires the presence of the agreement morpheme, then both number and gender agreement must surface as in the example in (48). In this example, there is a left dislocated element which triggers the presence of the agreement morpheme, which in *laísta* dialects surfaces as both gender and number agreement on the clitic: *las*.

(48) *A las balas las tiene miedo*
to the bullets$_{\text{FEM}}$ LAS has fear
'S/he is afraid of bullets.'

If the agreement morpheme is not present because it is not required in the structure, then what we expect is the non-agreeing *le* as in (49).

(49) *Le tiene miedo a las balas*
LE has fear to the bullets$_{\text{FEM}}$
'S/he is afraid of bullets.'

Note that under the current account we do not expect (50) to be possible. This is a good result because, as we have already seen, examples where agreement is

manifested in the clitic are ruled out when the doubling DP appears in its canonical position.

(50) ??*Las tiene miedo a las balas*
LAS has fear to the bullets$_{FEM}$
'S/he is afraid of bullets.'

In a way, we do not believe that the examples in (46) or (47) illustrate that there is no "la-for-las" since the contexts that could favor the lack of number agreement are also the contexts that require no agreement at all. In other words, claiming that (46) illustrates the lack of "la-for-las" assumes that *las* is possible to begin with, which is not right as illustrated by the unacceptability of (50).

5. Conclusions

In this paper, we have analyzed three previously unrelated phenomena: the presence of the accusative clitic, "le-for-les" and what we have labeled as "le-for-la." We have argued that the parallelism between these three constructions can be captured assuming that clitics come in three different flavors depending on their feature composition: accusative clitics are instances of agreement morphemes (as in Suñer's 1988 seminal work), and dative clitics may instantiate either purely an applicative morpheme (i.e. *le*) or a combination of an applicative morpheme and an agreement morpheme (i.e. *les* and *la/s*). We have shown that this analysis accounts for and explains the presence of the accusative clitic, the absence of number agreement with the dative in all dialects of Spanish, and the lack of both number and gender with the dative in *laísta* dialects of Spanish.

Furthermore, we believe that our account opens the door for a more unified treatment of clitics in Spanish. By proposing that dative clitics are composed of two different morphemes, an applicative morpheme and an agreement morpheme, and that the second one is similar (if not the same) to the accusative clitic, this approach opens the possibility of providing a uniform account to explain the differences and similarities between accusative and dative clitics.

Of course, there are a few questions that remain unanswered. For instance, what is the exact feature composition of dative clitics? Why does the agreement morpheme show up in the contexts that it does? Could the lack of agreement that we have covered be related to other cases? We hope that some of these questions will be answered in future research.

Acknowledgements

We would like to thank the audiences at the XLI Simposio de la Sociedad Española de Lingüística (Valencia, 2012), the 23rd Colloquium on Generative Grammar (Madrid, 2013) and the 44th Linguistic Symposium on Romance Languages (Ontario, 2014), as well as two anonymous reviewers, for comments and suggestions that have certainly improved our paper. Needless to say, we are responsible for the use we have made of their comments.

References

Alexiadou, Artemis. 2006. "Left dislocation (including CLLD)." In *The Blackwell Companion to Syntax*, ed. by M. Everaert and H. van Riemsdijk, 668–699. Malden: Blackwell. doi:10.1002/9780470996591.ch37

Bleam, Tonia. 1999. *Leísta Spanish and the Syntax of Clitic Doubling*. University of Delaware, doctoral dissertation.

Boeckx, Cedric, and Youngmi Jeong. 2004. "The fine structure of intervention in syntax." In Chungja Kwon and Wonbin Lee (eds), *Issues in Current Linguistic Theory: A Festschrift for Hong Bae Lee*, 83–116.

Casares, Julio. 1918. *Crítica efímera*. Madrid: Saturnino Calleja S.A.

Cuervo, María Cristina. 2003. *Datives at Large*, Ph.D. Thesis. Cambridge, Mass.: MIT.

Cuervo, María Cristina. 2010. "Against Ditransitivity." *Probus* 22 (2): 151–180. doi:10.1515/prbs.2010.006

Cuervo, Rufino José. 1907/1955. *Apuntaciones críticas sobre el lenguaje bogotano, con frecuente referencia al de los países de Hispano-América* (9th ed.). Bogotá: Instituto Caro y Cuervo.

DeMello, George. 1992. "'Le' for 'les' in the spoken educated Spanish of eleven cities." *Canadian Journal of Linguistics* 37 (4): 407–430.

Demonte, Violeta. 1995. "Dative alternation in Spanish." *Probus* 7 (1): 5–30. doi:10.1515/prbs.1995.7.1.5

Fernández-Ordóñez, Inés. 1994. "Isoglosas internas del castellano: El sistema referencial del pronombre átono de tercera persona." *Revista de Filología Española* LXXIV (1): 71–125. doi:10.3989/rfe.1994.v74.i1/2.499

Fernández-Soriano, Olga. 1999. "El pronombre personal. Formas y distribuciones. Pronombres átonos y tónicos." In *Gramática descriptiva de la lengua española*, ed. by Ignacio Bosque and Violeta Demonte, 1209–1274. Madrid: Espasa Calpe.

Huerta Flores, Norohella. 2005. "Gramaticalización y concordancia objetiva en el español: Despronominalización del clítico dativo plural." *Verba* 32: 165–190.

Martín, Francisco Jesús. 2012. *Deconstructing Catalan Objects*, Ph.D. Thesis. New York: New York University.

Ormazabal, Javier, and Juan Romero. 2013. "Object Clitics, Agreement and Dialectal Variation." *Probus: International Journal of Latin and Romance Linguistics* 25 (2): 301–344.

RAE. 1885. *Gramática de la lengua castellana*. Madrid: G. Hernando.

RAE & AALE. 2009. *Nueva gramática de la lengua española*. Madrid: Espasa Libros.

Rinni, Joel. 1988. "A New Perspective on the Origin of le for les." *Journal of Hispanic Philology* 12 (3): 207–219.

Romero, Juan. 2012. "Accusative Feminine Datives in Spanish." In *Variation in Datives: A Microcomparative Perspective*, ed. by Beatriz Fernández and Ricardo Etxepare, 283–300. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199937363.003.0010

Sportiche, Dominique. 1996. "Clitic Constructions." In *Phrase Structure and the Lexicon*, ed. by J. Rooryck and L. Zaring, 213–276. Dordrecht: Kluwer. doi: 10.1007/978-94-015-8617-7_9

Sturgis, Cony. 1927. "Uso de le por les." *Hispania* 10 (4): 251–254. doi: 10.2307/331165

Suñer, Margarita. 1988. "The Role of Agreement in Clitic Doubled Constructions." *Natural Language and Linguistic Theory* 6 (3): 391–434. doi: 10.1007/BF00133904

Uriagereka, Juan. 1988. *On Government*. University of Connecticut, doctoral dissertation.

CHAPTER 6

The morphological markedness of φ

Evidence from perfective auxiliaries in Southern Italian dialects

Giuseppe Torcolacci
Leiden University

This paper investigates the morphological markedness of φ features encoded on perfective auxiliaries in a subset of Upper Southern Italian dialects (USIDs).

Following a Distributed Morphology approach (cf. Halle & Marantz 1993, 1994, a.o.), we treat perfective auxiliaries in USIDs as syntactic heads composed of Tense and φ features whose phonological entries are inserted at PF by means of a process called Spell-Out.

Here, we argue that the overt marking of φ features attested on perfective auxiliaries in a group of USIDs depends on the application of a post-syntactic operation called *Default Marking*, according to which a dedicated set of φ features gets overtly marked only if its grade of markedness is uniform with the one expressed by Tense.

Keywords: Southern Italian dialects, auxiliary selection, Raddoppiamento Fonosintattico, metaphony, markedness, default, tense, φ-features

1. Introduction

The gist of this paper is to investigate the morphological markedness of φ features expressed on perfective auxiliaries in a subset of Upper Southern Italian dialects (USIDs). The dialects under investigation correspond to those spoken in the geolinguistic area that stretches from northern Apulia / central Campania up to northern Calabria. Map 1 illustrates the geographic extension of this dialectal area, which from now on we will refer to as Central Southern Italian dialects (CSIDs).

As far as the overt marking of φ features on present perfect auxiliaries is concerned, a large number of CSIDs behave differently from the rest of USIDs. Many CISDs generally disallow the overt marking of 2nd person singular on a present

DOI 10.1075/rllt.11.06tor

Map 1. The geolinguistic extension of USIDs and CSIDs[1]

perfect auxiliary.[2] The overt marking of 2nd person singular on a present perfect auxiliary is instead common to all other USIDs. Furthermore, CSIDs, different from

1. Map 1, taken from google, marks the isoglosses that delimit the geographic extension of USIDs and CSIDs. The drawn isoglosses are based on the Carta dei Dialetti d'Italia (cf. Pellegrini 1977).

2. A handful of dialects spoken in Campania and Apulia, as well as a large number of southern Lucanian dialects, display the presence of inflected forms for 2nd person singular present perfect auxiliaries. The dialect of Polignano a Mare [Apulo-Barese], for instance, admits the overt expression of 2nd person singular by selecting the root /ɛ/, which is not selected in the case of 1st and 3rd person singular present perfect forms (cf. /aɟɟ/ 'have-PRES-1SG' and /a/ 'have-PRES-3SG'). In southern Lucanian dialects, 2nd person singular present perfect auxiliaries share the same root with those expressing 1st and 3rd person singular. In these dialects, the overt marking of 2nd person singular is expressed by means of a dedicated φ marker occurring in word-final position (cf. Senise [Southern Lucano] ædd͡ʒə 'have-PRES-1SG' / æ**i** 'have-PRES-2SG' / æ 'have-PRES-3SG' (Manzini & Savoia 2005, II)).

the rest of USIDs, allow the presence of *Raddoppiamento Fonosintattico*, henceforth RF, after a 3rd person singular present perfect auxiliary.[3]

These facts are exemplified in the paradigms in (1) and (2). (1) corresponds to the CSID of Mola di Bari, whereas (2) corresponds to the dialect of San Benedetto del Tronto, spoken outside the area of CSIDs.[4]

(1) Mola di Bari (Apulo-Barese)

aʃʃ	*ˈfatt/parˈlə:t/parˈtʉ:t*	'have-PRES-1SG done/spoken/left'
a	*ˈfatt/parˈlə:t/parˈtʉ:t*	'have-PRES-2SG done/spoken/left'
(')a	***ff****att/****pp****arˈlə:t/****pp****arˈtʉ:t*	'have-PRES-3SG done/spoken/left'
am	*ˈfatt/parˈlə:t/parˈtʉ:t*	'have-PRES-1PL done/spoken/left'
avet	*ˈfatt/parˈlə:t/parˈtʉ:t*	'have-PRES-2PL done/spoken/left'
an	*ˈfatt/parˈlə:t/parˈtʉ:t*	'have-PRES-3PL done/spoken/left'

(2) San Benedetto del Tronto (Southern Marchigiano)

so	*ˈviʃtə/dərˈmitə/veˈnu:tə*	'be-PRES-1SG seen/slept/come'
si	*ˈviʃtə/dərˈmitə/veˈnu:tə*	'be-PRES-2SG seen/slept/come'
a	*ˈviʃtə/dərˈmitə/veˈnu:tə*	'have-PRES-3 seen/slept/come'
ʃemə	*ˈviʃtə/dərˈmitə/veˈnu:tə*	'be-PRES-1PL seen/slept/come'
ʃetə	*ˈviʃtə/dərˈmitə/veˈnu:tə*	'be-PRES-2PL seen/slept/come'

(Manzini & Savoia 2005, II: 682–683)

3. *Raddoppiamento Fonosintattico*, or RF, consists of the gemination of the first consonant of Word2 in the context Word1#Word2 (Chierchia 1986; Fanciullo 1986, 1997; Bertinetto & Loporcaro 1988; Loporcaro 1988, 1997a, 1997b; Vincent 1988; Nespor 1993; Waltereit 2004; Passino 2012; Torcolacci 2014, 2015, a.o.). In the traditional literature, RF is considered to result from regressive consonantal assimilation applying at external sandhi sites that is believed to have taken place in the period of transition from Latin to southern Italo-Romance (cf. Schuchardt 1874; Hall 1964; Loporcaro 1997b, a.o.). As an example, RF after conjunction *e* is considered to derive from the assimilation of the consonant *t* in word-final position of the Latin word ET with the first consonant of the following word (cf. ET VIDET [Latin] > e **vv**ede [Southern Italian dialects]). RF is not found only in Southern Italian dialects, but its presence is attested also for Italian dialects spoken in the central part of the peninsula, Sardinian and Corsican included.

4. The dialect of Mola di Bari in (1), different from the dialect of San Benedetto del Tronto in (2), opts for a Spanish-like system of auxiliary selection, whereby HAVE is the present perfect auxiliary selected for all persons in the paradigm. This pattern is not limited to the variety of Mola di Bari in (1), but also found in other CSIDs. Differently from CSIDs, the rest of USIDs displays a different type of auxiliary selection in the present perfect. In these dialects, BE is generally selected when the subject is 1st and 2nd person, whereas HAVE is generally chosen in correspondence with a 3rd person subject (cf. Cocchi 1995; D'Alessandro & Roberts 2010; Ledgeway 2000; Manzini & Savoia 2005; Legendre 2010; Loporcaro 2010, a.o.). The pattern of auxiliary selection observed for the dialects of Mola di Bari and San Benedetto del Tronto in (1) and (2), respectively, differs from the one attested for Italian and French where auxiliary selection is claimed to be dependent on the verbal class, or *Aktionsart*, of the past participle the auxiliary merges with (cf. Perlmutter 1978; Burzio 1986; Hubert & Rindler-Schjerve 1987; Chierchia 1989; Legendre 1989; van Valin 1990; Loporcaro 1998; Sorace 2000, a.o.).

The marking of 2nd person singular on a present perfect auxiliary is not restricted to the dialect of San Benedetto del Tronto, but rather attested for other (and not only) Romance languages. In Italian and Romanian, for instance, 2nd person singular is overtly expressed by means of /i/ that occurs in word-final position (*(h)ai*, 'have-PRES-2SG'). In Spanish and a subset of Northern Italian dialects, the consonant that encodes 2nd person singular is /s/ (*(h)as*, 'have-PRES-2SG') appearing in word-final position.

Despite the plural auxiliaries in (1) and (2) being all marked for their φ reference, the ones in the singular paradigm show that:

(3) a. 2nd person is overtly encoded in (2), and not in (1);
b. 3rd person singular HAVE triggers RF in (1), and not in (2);
c. 1st person in (1) and (2) is overtly marked by means of a dedicated φ marker.

Following a Distributed Morphology, or DM, approach (cf. Halle & Marantz 1993, 1994; Calabrese 1994; Harley 1994; Harris 1994; Embick 1995; Noyer 1997, a.o.), we will consider perfective auxiliaries as syntactic heads composed of a bundle of abstract morphosyntactic features. The insertion of phonological exponents replacing the content of features contained in the auxiliaries applies at PF by means of a mechanism called Spell-Out.

The marking strategies of φ observed in (1), different from the ones in (2), will be claimed to derive from the application of a post-syntactic operation called *Default Marking*, according to which 2nd person singular never gets overtly marked on a present perfect auxiliary in a group of CSIDs. In our model, *Default Marking* is a morphological operation active in the module of the grammar sandwiched between syntax and PF.

This paper is structured as follows: in §2, the typology of morphological markedness of φ features attested on perfective auxiliaries in CSIDs will be examined. §3 will consider the syntactic nature of perfective auxiliaries in CSIDs. §4 will present the post-syntactic operation of *Default Marking* active in the case of perfective auxiliaries in CSIDs. §5 summarizes and concludes the paper.

2. The typology of morphological markedness of φ on perfective auxiliaries in CSIDs

Forchheimer (1953: 6) claims that languages tend to exhibit a mismatch in the morphological marking between 1st/2nd and 3rd person agreement markers. Based on a cross-linguistic observation, Forchheimer claims that verbs generally mark 1st and 2nd person by means of a dedicated φ marker, whereas the overt marking

of 3rd person, on the other hand, is not commonly attested. This behavior is attested for almost all Romance languages, whereby 1st and 2nd person agreement markers are always overtly expressed, with the exclusion of 3rd person ones. These facts have been illustrated above with reference to the dialect of San Benedetto del Tronto in (2).

CSIDs, different from the rest of USIDs, opt for a different strategy of φ marking in the case of perfective auxiliaries. In these languages, a marker expressing 2nd person never gets overtly expressed on a present perfect auxiliary. Furthermore, present perfect auxiliaries expressing 3rd person singular always license RF. RF is found only in this context and banned elsewhere. These facts are illustrated in (4) and (5) with reference to the CSIDs of Mola di Bari and Airola, respectively.

(4) Mola di Bari (Apulo-Barese)

aʝʝ	*ˈfatt/parˈlə:t/parˈtʉ:t*	'have-PRES-1SG done/spoken/left'
a	*ˈfatt/parˈlə:t/parˈtʉ:t*	'have-PRES-2SG done/spoken/left'
(ˈ)a	***f****fatt/****p****parˈlə:t/****p****parˈtʉ:t*	'have-PRES-3SG done/spoken/left'
am	*ˈfatt/parˈlə:t/parˈtʉ:t*	'have-PRES-1PL done/spoken/left'
avet	*ˈfatt/parˈlə:t/parˈtʉ:t*	'have-PRES-2PL done/spoken/left'
an	*ˈfatt/parˈlə:t/parˈtʉ:t*	'have-PRES-3PL done/spoken/left'

(5) Airola (Central Campano)

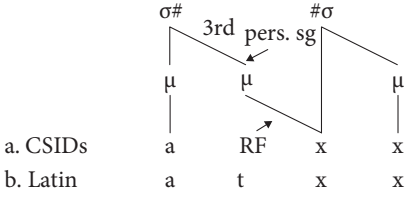
aʝʝ	*ˈvistə/ˈfattə*	'have-PRES-1SG seen/done'
a	*ˈvistə/ˈfattə*	'have-PRES-2SG seen/done'
a	***v****ˈvistə/****f****fattə*	'have-PRES-3SG seen/done'
ammu	*ˈvistə/ˈfattə*	'have-PRES-1PL seen/done'
atə	*ˈvistə/ˈfattə*	'have-PRES-2PL seen/done'
annə	*ˈvistə/ˈfattə*	'have-PRES-3PL seen/done'

We claim that RF triggered by 3rd person singular HAVE in CSIDs corresponds to the marker of 3rd person singular (cf. Torcolacci 2014, 2015).[5] These facts are illustrated by means of the phonological structure in (6) which is based on the Moraic Theory by Hayes (1989).[6]

5. The dialects of Mola di Bari and Airola do not feature RF in the case of 3rd person singular lexical verbs. In these dialects, RF is triggered only by 3rd person singular HAVE and unattested with other verbal forms (see Torcolacci 2014, 2015).

6. According to Hayes (1989), phonological representations are composed of three layers: the syllable (σ), the mora (μ), and the segment (x).

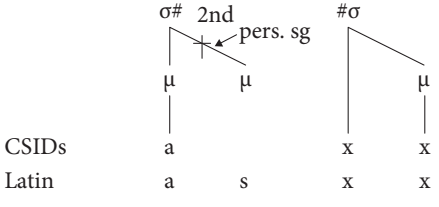
(6) 3rd person sg pres. perfect HAVE: *a* + RF < Lat. *HA(BE)T



Latin 3rd person singular HAVE hosted the 3rd person singular agreement marker *t* in word-final position, which at a certain point in diachrony, was deleted in CSIDs. The deletion of the segment *t*, however, did not provoke the deletion of the mora associated to it, which stayed alive in the phonological structure, and required the regressive spreading of the next consonant in the linear string, thus, provoking RF. In these terms, we claim that the empty mora triggering RF is endowed with 3rd person singular specification (cf. Torcolacci 2014, 2015).

Different from (6), 2nd person singular present perfect HAVE in CSIDs does not trigger RF despite the fact that Latin 2nd person singular HAVE hosted the consonant *s* in word-final position (cf. (7)).

(7) 2nd person sg pres. perfect HAVE: *a* + Ø < Lat. *HA(BE)S



The fall of consonant *s* in word-final position of *HA(BE)S also triggered the deletion of the mora associated to it in the period of transition from Latin to Southern Italo-Romance, and for this reason RF is not attested. To sum up, the difference between 2nd and 3rd person singular present perfect HAVE in CSIDs is that 2nd person HAVE allowed the deletion of the segment *s* and of the mora associated to it, while 3rd person HAVE did not undergo the same process and only the segment *t*, and not the mora, disappeared from the phonological representation.

The overt marking of φ features in the case of pluperfect auxiliaries in CSIDs, however, strongly differs from the one observed in (4) and (5). While 2nd person singular present perfect HAVE in (4) and (5) is bare, thus, not hosting any φ marker, 2nd person singular pluperfect HAVE in (8) and (9) overtly expresses 2nd person

singular. The overt marking of 2nd person singular on a pluperfect auxiliary is expressed by means of metaphony.[7]

(8) Mola di Bari (Apulo-Barese)

aˈvɛv	*manˈdʒət/vəˈvɣʷt*	'have-PAST-1SG eaten/drunk'
*aˈ**vi**v*	*manˈdʒət/vəˈvɣʷt*	'have-PAST-2SG eaten/drunk'
aˈvɛv	*manˈdʒət/vəˈvɣʷt*	'have-PAST-3SG eaten/drunk'
*aˈ**vemm***	*manˈdʒət/vəˈvɣʷt*	'have-PAST-1PL eaten/drunk'
*aˈ**vivər***	*manˈdʒət/vəˈvɣʷt*	'have-PAST-2PL eaten/drunk'
*aˈ**vevən***	*manˈdʒət/vəˈvɣʷt*	'have-PAST-3PL eaten/drunk'

(9) Airola (Central Campano)

aˈle:və	*ˈfattə/ˈvistə*	'have-PAST-1SG done/seen'
*aˈl**i:**və*	*ˈfattə/ˈvistə*	'have-PAST-2SG done/seen'
aˈle:və	*ˈfattə/ˈvistə*	'have-PAST-3SG done/seen'
*aˈle:və**mə***	*ˈfattə/ˈvistə*	'have-PAST-1PL done/seen'
*aˈle:və**və***	*ˈfattə/ˈvistə*	'have-PAST-2PL done/seen'
*aˈle:və**nə***	*ˈfattə/ˈvistə*	'have-PAST-3PL done/seen'

In the traditional literature, metaphony in the case of 2nd person singular HAVE in CSIDs is treated as been triggered by the presence of a high vowel in word-final position attested in the older stages of the languages. In this sense, the presence of a high vowel on 2nd person singular HAVE in (8) and (9) is understood to be the fossil of metaphony that was active in diachrony. If this were true, however, metaphony should be found also on 2nd person singular lexical verbs in CSIDs. Contrary to what is expected, a subset of 2nd person singular lexical verbs of the dialect of Mola di Bari does not feature metaphony, as shown in (10).

(10) Mola di Bari
ˈmandʒ / ˈparl
'eat / speak-PRES-2SG'

The absence of a high stressed vowel on the 2nd person singular verbs in (10) suggests that metaphony should be understood as a phonological phenomenon that is active only in some contexts in CSIDs, i.e. with pluperfect auxiliaries and not

7. The term metaphony refers to a phonological process whereby a high vowel has a raising influence on a preceding stressed vowel when this bears a mid or low feature. In a large number of Italo-Romance dialects, low or mid vowels can be raised, or diphthongized, when preceding a high vowel occurring in final position: *vérde* ('green-SG.') versus *virdi* ('green-PL.'), *pédi* ('foot') versus *pjédi* ('feet') (cf. Calabrese 2011). *Virdi* hosts a high vowel in last position, whose presence licenses the stressed mid-low vowel *é* to be raised to *i*. The vowel *i* in *pjédi*, on the other hand, does not trigger vowel raising but a diphthong in stressed position.

with a subset of lexical verbs, whose function is that of expressing 2nd person.[8] The overt marking of 2nd person by means of metaphony is not restricted to the singular paradigm. In the plural paradigm in (8), indeed, the stressed high vowel /i/ is attested only on 2nd plural pluperfect HAVE, and excluded elsewhere.

Table 1 summarizes the marking strategies of φ in the case of present perfect and pluperfect auxiliaries in the dialects in (4)–(5) and (8)–(9). There, the overt marking of φ features in the singular paradigm is illustrated. + indicates where the overt marking of a given φ feature is at play. –, conversely, indicates that no overt marking for a given feature is attested.

Table 1. The marking of φ on perfective southern Italian auxiliaries

	Present perfect auxiliary			**Pluperfect auxiliary**		
	1st pers.	**2nd pers.**	**3rd pers.**	**1st pers.**	**2nd pers.**	**3rd pers.**
Mola d. B. (cf. 4 & 8)	+	–	+	–	+	–
Airola (cf. 5 & 9)	+	–	+	–	+	–

In Table 1, it can be observed that the dialects of Mola di Bari and Airola allow the overt marking of 1st and 3rd person singular when expressed on a present perfect auxiliary. The overt marking of 2nd person singular on a present perfect auxiliary, instead, is never obtained. On the other hand, the overt marking of 1st and 3rd person singular is never found on a pluperfect auxiliary. Pluperfect auxiliaries in CSIDs only allow the overt marking of 2nd person singular.

3. Auxiliaries in CSIDs

3.1 The syntactic nature

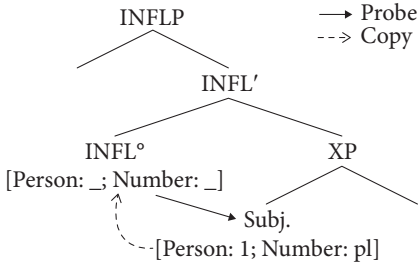
We consider perfective auxiliaries in CSIDs as syntactic heads directly merged in INFL. Based on Chomsky (1995, 1999, 2000, 2001), we argue that auxiliaries, in the same fashion as lexical verbs, are endowed with uninterpretable Person and Number features. In order for these features to be interpreted, the operation *Agree* between the DP-subject, e.g. the goal, and the auxiliary, e.g. the probe, must take place.[9] *Agree* predicts that the interpretable values on a goal are copied by the cor-

8. The reason why RF is found on a 2nd person singular pluperfect auxiliary, and not on a subset of lexical verbs in CSIDs, is explained in Torcolacci (2015).

9. According to Chomsky (1999), *Agree* between a probe and a goal is possible only if the probe c-commands the goal and both the probe and the goal are active, thus, endowed with a Case feature.

responding uninterpretable ones encoded on the probe. After *Agree* takes place, the Person and Number features in INFL are fully specified.

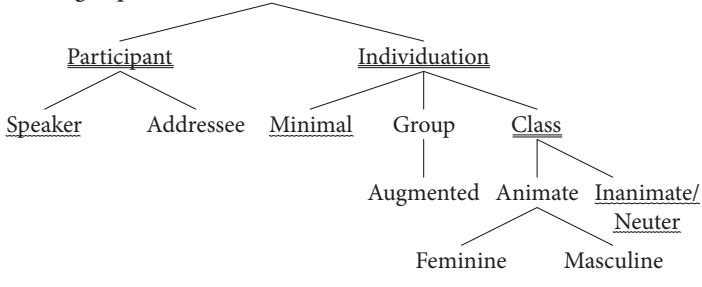
(11)



(11) indicates that the DP-subject expresses 1st person and plural values. These values are copied by the Person and Number features in INFL, and, therefore, INFL is also interpretable for the same values expressed on the subject.

According to Harley and Ritter (2002), morphosyntactic φ features expressed on pronouns are structured within a geometry. This is to say that features are monovalent and organized within a hierarchical structure (cf. (12)).

(12) **Referring Expression** (=Pronoun)



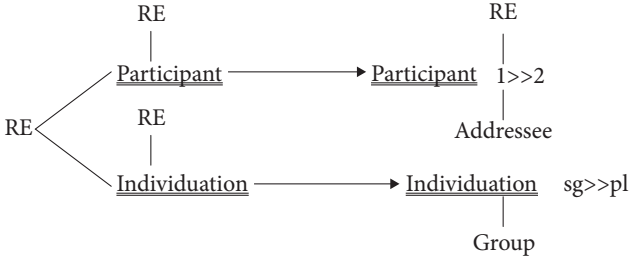
(Harley & Ritter 2002:8)

The features Participant and Individuation in (12) refer to Person and Number, respectively. These two features are further branched into sub-nodes. Participant branches into Speaker and Addressee. The former expresses information for 1st person singular while the latter expresses 2nd person singular. Individuation, on the other hand, has three sub-nodes, amongst which Minimal, Group and Class are included. Minimal refers to singular and Group to plural.[10] The sub-nodes of Participant and Individuation bear a different type of markedness. Speaker and Minimal, which are curly underlined in the geometry in (12), are the default

10. We leave aside here the discussion referring to Class. This feature corresponds to Gender, which, in turn, is branched into Neuter/Inanimate, Masculine and Feminine (see Harley & Ritter 2002).

morphosyntactic features within Participant and Individuation, respectively. Addressee and Group, conversely, are marked (cf. Harley & Ritter 2002). The reason why Speaker and Minimal are considered as defaults derives from some observations related to the acquisition of pronouns. Speaker is always acquired before Addressee and Minimal is always learnt before Group. This means that at the early stage of acquisition of pronouns, the default interpretation for Participant coincides with Speaker, while the default interpretation for Individuation corresponds to Minimal. Addressee and Group, in being acquired later, are considered to be marked and their interpretation never coincides with that of their mother nodes (cf. (13)).

(13)



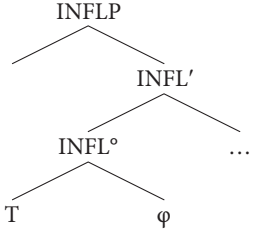
(Harley & Ritter 2002:28)

An important fact that needs to be taken into account is that the feature geometry in (12) refers to pronoun, and not, for instance, to agreement markers. Indeed, as the Harley and Ritter (2002) point out, "several major research questions now arise: the nature of the relationship between the geometry and the syntactic component, in particular with respect to agreement phenomena" (cf. Harley & Ritter 2002: 53).

Here, we propose that the feature geometry in (12) also holds for agreement markers. More specifically, we argue that morphosyntactic φ features encoded on perfective auxiliaries in CSIDs are structured in the same way as (12).

Besides φ, we claim that perfective auxiliaries in CSIDs are endowed with the feature Tense (T). These facts are summarized by means of the structure in (14).

(14)



Based on Ritter & Wiltschko (2009, 2010), we consider INFL as a syntactic head endowed with an universal [*u*coin] feature. The [*u*coin], or [*u*coincidence], feature is underspecified and its value depends on the anchoring mechanism pertaining between the event and the utterance situation. In Indo-European languages, [*u*coin] is a feature that expresses whether the event and the utterance situation coincide in time.[11] For this reason, we claim that [*u*coin] is a feature encoded in Tense. If the event situation, encoded in Spec,VP, coincides in time with the utterance situation, expressed in Spec,INFLP, then [*u*coin] expresses a + value. On the other hand, if the event and the utterance situation do not converge in their time reference, then [*u*coin] expresses a – value. Let us now consider the sentences in (15).

(15) Mola di Bari (Apulo-Barese)

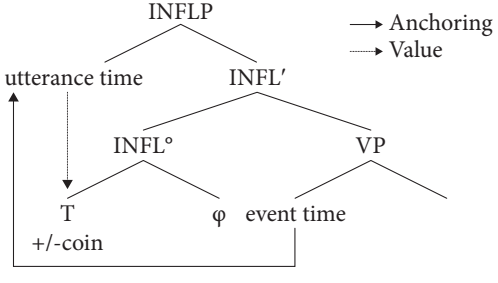
a. *aʃʃ* *ˈfatt* *na* *ˈtɔːrtə*
have.PRES.1SG done a.FM.SG pie
'I have made a pie.'

b. *aˈvɛv* *ˈfatt* *na* *ˈtɔːrtə*
have.PAST.1SG done a.FM.SG pie
'I had made a pie.'

In (15a), the event of making a pie has direct consequences to the utterance situation and in this case the present form of auxiliary HAVE is chosen. On the other hand, the event of making a pie in (15b) does have direct consequences on the utterance situation, but on a moment prior to that. There, the past form of auxiliary HAVE is selected.

The structure in (16) shows how the mechanism of valuation of [*u*coin] works in the case of the present perfect and pluperfect auxiliaries in (15).

(16)



We consider the value + expressed on [*u*coin] as a default. Conversely, the value – is treated as marked. This assumption relies on the idea put forward by Holmberg

11. Ritter and Wiltschko (2009, 2010) claim that INFL is not only composed of the grammatical categories Tense and Person (Person corresponds to φ in our account). The authors claim that INFL also includes the category Location.

and Roberts (2010), discussed in the following section, according to which features endowed with the same type of specification are considered to license unmarked, or default, syntactic configurations.

As the data in (4)–(5) and (8)–(9) indicate (cf. §2), the information expressed by [*u*coin] seems to determine the type of morphosyntactic φ features to be overtly spelled-out on the auxiliary. If [*u*coin] is valued as +, then only Speaker and Minimal get overtly marked. On the other hand, if [*u*coin] is valued as −, then only Addressee is marked at PF. In the next section, it will be investigated why the spell-out of morphosyntactic φ features encoded on perfective auxiliaries in CSIDs is sensitive to the information expressed by [*u*coin].

4. The morphophonological markedness of φ on perfective auxiliaries in CSIDs: The operation of *Default Marking*

In this part, we argue that the overt marking of morphosyntactic φ features encoded on perfective auxiliaries in CSIDs depends on the application of a post-syntactic operation called *Default Marking*. §4.1. will consider the uniformity of markedness of features encoded in INFL as a requirement for the *Default Marking* to apply. §4.2. will consider in detail the post-syntactic operation of *Default Marking*.

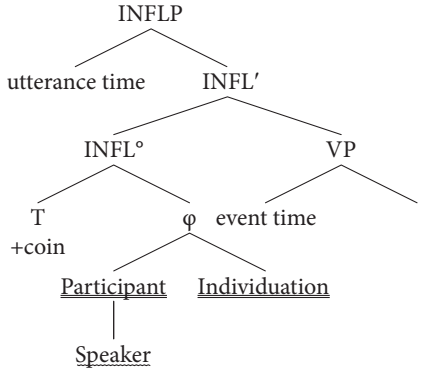
4.1 The uniformity of markedness in INFL

Holmberg and Roberts (2010) claim that markedness in syntax can be understood in terms of uniformity of values expressed by features. With reference to the EPP feature, Holmberg and Roberts propose that if all heads endowed with an EPP feature in a syntactic structure express all the same value, then, either a + or a − value, then an unmarked, (i.e. default), syntactic configuration is obtained. In the former case, namely in the presence of [+EPP] for all syntactic heads endowed with movement triggering properties, a harmonic head-final syntactic configuration is obtained. In the latter case, namely when all syntactic heads able to trigger movement are specified for [−EPP], a harmonic head-initial syntactic configuration is attested. The presence of [+EPP] for some heads and [−EPP] for some others would give rise to mixed configurations, which, according to Holmberg and Roberts (2010), are marked.

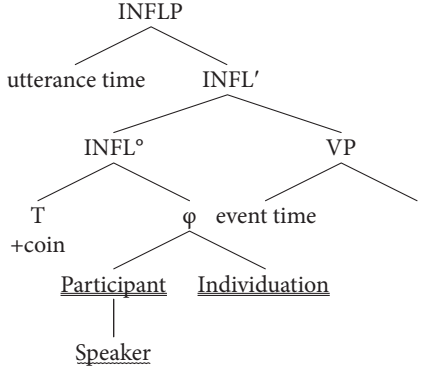
Based on this markedness convention, we propose that an unmarked, namely default, configuration is obtained also when all features contained in a syntactic head are specified for the same value. This is to say that, with reference to perfective auxiliaries in CSIDs, a default configuration is obtained if both Tense and φ are

uniform in their grade of markedness. In (17a), Tense and φ bear the same type of markedness and an unmarked configuration is obtained. In (17b), on the other hand, Tense and φ are not uniform in their grade of markedness, and, therefore, a marked configuration is at play.

(17) a.



b.



In (17a), φ is specified for Speaker and Individuation, and no marked values branching below Participant and Individuation are found. For this reason, we consider the grammatical category φ in (17a) as expressing an unmarked value which is uniform in its grade of markedness with [+coin]. In (17b), instead, [*u*coin] is specified as –. The value expressed on [–coin] does not coincide in its grade of markedness with the one expressed by φ, and, therefore, a marked configuration is obtained.

4.2 The Default Marking

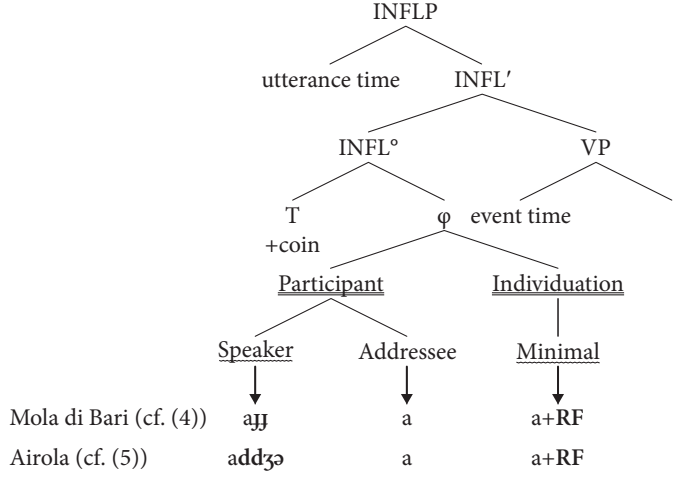
Here, we claim that morphosyntactic φ features encoded on perfective auxiliaries in a large number of CSIDs are affected by the post-syntactic operation of *Default*

Marking, that is supposed to be active in the morphological component of the grammar. The definition of *Default Marking* is given in (18).

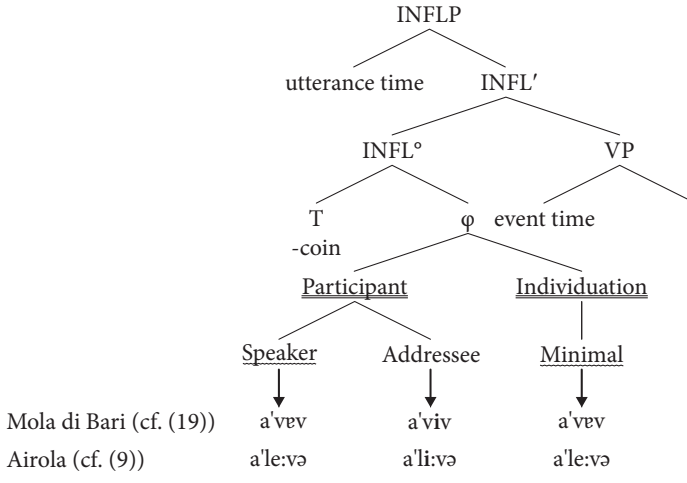
(18) *Default Marking*
The morphological marking of a φ feature can only take place if all features bear the same markedness on the functional head that hosts them.
(Torcolacci 2015: 123)

Based on the *Default Marking* operation in (18), we claim that default morphosyntactic φ features encoded on perfective auxiliaries in CSIDs get overtly marked if [*u*coin] bears a + value, which also corresponds to a default. On the other hand, if [*u*coin] bears a − value, which is marked, then only marked morphosyntactic features get overtly expressed. These facts are represented by means of the diagrams in (19) and (20). In (19), we observe that only Speaker and Minimal are overtly marked, with the exclusion of Addressee. There, [*u*coin] is valued as +. In (20), on the other hand, Addressee is the only feature that gets overtly marked. This is due to the fact that Addressee combines with [−coin]. In (19) and (20), reference is made to the dialects of Mola di Bari and Airola presented in (4)–(5) and (8)–(9).

(19)



(20)



In (19), Speaker and Minimal are overtly marked because they express the same type of markedness with [+ coin]. For this reason, *Default Marking* can apply. Addressee, on the other hand, is a marked morphosyntactic φ feature that expresses a different type of markedness from [+ coin]. In this case, *Default Marking* cannot apply.

The overt marking of Speaker is obtained by means of a morphological marker realized at final position of the auxiliary. In Mola di Bari, the marker expressing Speaker corresponds to the double fricative /ʝ/, while in Airola the marker expressing Speaker corresponds to the affricate /ddʒ/. In both dialects, the overt marking of Minimal is instead obtained by means of RF, which, as argued in §2, corresponds to the way 3rd person singular, i.e. Minimal, gets morpho-phonologically marked on 3rd person singular present perfect HAVE.

In (20), Addressee is overtly marked because it shares the same grade of markedness with [−coin]. In this case, *Default Marking* can apply. Speaker and Minimal, which are default morphosyntactic φ features, are not uniform in their grade of markedness with [−coin]. In this case, *Default Marking* cannot be obtained and as a consequence Speaker and Minimal are morphologically unmarked.

The overt marking of Addressee is obtained by means of metaphony. The 1st and 3rd person singular form of HAVE expressing past information, instead, are syncretic and no special φ-marking strategy is obtained there.

5. Summary and conclusions

In this paper, we have considered the morphological markedness of morphosyntactic φ features encoded on perfective auxiliaries in CSIDs.

A subset of CSIDs, different from other USIDs, are special in the way they overtly mark φ features on perfective auxiliaries. In these dialects, the overt marking of φ features on perfective auxiliaries seems to depend on the information expressed by Tense: if Tense is Present, then Speaker and Minimal get overtly realized, with the exclusion of Addressee. On the other hand, if Tense is Past, then Addressee is overtly expressed and Speaker and Minimal do not get overtly marked. Speaker and Minimal refer to 1st and 3rd person singular, whereas Addressee refers to 2nd person singular (cf. Harley & Ritter 2002).

Following Harley and Ritter (2002), we have considered morphosyntactic φ features as being organized within a feature geometric representation. Moreover, morphosyntactic φ features have been argued to inherit a different grade of markedness, according to the way they are acquired by children. Speaker and Minimal are the first features to be acquired, and, therefore, they are considered unmarked, or default. On the other hand, Addressee, which is learnt after Speaker and Minimal, is marked.

We have considered perfective auxiliaries in CSIDs to be syntactic objects directly merged in INFL and composed of the features Tense and φ. Based on Ritter and Wiltschko (2009, 2010), we have considered Tense to be endowed with an [*u*coin] feature, whose function is to anchor the event time, encoded in Spec,VP, and the utterance time, in Spec,INFL. If the event and the utterance time coincide, then [*u*coin] is valued as +, which, in our model, is an unmarked (i.e. default) value. Conversely, if the event and the utterance time do not coincide, then [*u*coin] is valued as −, which we consider to be marked.

We have argued that the morphological markedness of φ on perfective auxiliaries in CSIDs depends on the application of a post-syntactic operation called *Default Marking*. *Default Marking* says that morphosyntactic φ features get overtly spelled-out if their grade of markedness coincides with the one expressed by [*u*coin]. This is to say that if [+ coin] combines with either Speaker or Minimal, then Speaker or Minimal gets overtly marked. On the other hand, if [−coin] combines with Addressee, then Addressee gets overtly marked.

Default Marking, which in our framework is considered to apply in the morphological component of the grammar, explains why Addressee does not get overtly marked when [*u*coin] is valued as +, as well as why Speaker and Minimal are left morphologically unmarked in the case of [−coin]. In both cases, [*u*coin] and φ express different grades of markedness and for this reason *Default Marking* cannot apply.

References

Bertinetto, Pier M., and Michele Loporcaro. 1988. *Certamen phonologicum. Papers from the 1987 Cortona phonology meeting*. Torino: Rosenberg & Sellier.

Burzio, Luigi. 1986. *Italian Syntax*. Dordrecht: Reidel. doi:10.1007/978-94-009-4522-7

Calabrese, Andrea. 1994. "Syncretism phenomenon in the clitic systems of Italian and Sardinian dialects and the notion of morphological change." *Proceedings of NELS* 25: 151–174.

Calabrese, Andrea. 2011. "Metaphony in Romance." *The Blackwell Companion to Phonology*: 2631–2661.

Chierchia, Gennaro. 1986. "Length, syllabification and the phonological cycle in Italian." *Journal of Italian Linguistics* 8: 5–33.

Chierchia, Gennaro. 1989. "A semantics for unaccusatives and its syntactic consequences." Ms., Cornell University, Ithaca, N.Y.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1999/2001. "Derivation by Phase." In Ken Hale: *A Life in Language*, ed. by Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. "Minimalist Inquiries: the Framework." In *Step by step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, ed. by Roger Martin, David Michaels and Juan Uriagereka, 89–156. Cambridge, MA: MIT Press.

Cocchi, Gloria. 1995. *La selezione dell'ausiliare*. Padova: Unipress.

D'Alessandro, Roberta, and Ian Roberts. 2010. "Past participle agreement in Abruzzese: split auxiliary selection and the null-subject parameter." *Natural Language and Linguistic Theory* 28, 41–72. doi:10.1007/s11049-009-9085-1

Embick, David. 1995. "Mobile inflections in Polish." *Proceedings of NELS* 25: 127–142.

Fanciullo, Franco. 1986. "Syntactic reduplication and the Italian dialects of the Centre-South." *Journal of Italian Linguistics* 8: 67–104.

Fanciullo, Franco. 1997. *Raddoppiamento Sintattico e la ricostruzione linguistica del Sud italiano*. Pisa: ETS.

Forchheimer, Paul. 1953. *The category of person in language*. Berlin: Walter de Gruyter. doi:10.1515/9783111562704

Hall, Robert. 1964. "Initial consonants and syntactic doubling in West Romance." *Language* 40: 551–556.

Halle, Morris, and Alec Marantz. 1993. "Distributed Morphology and the Pieces of Inflection." *The View from Building* 20: 111–176.

Halle, Morris, and Alec Marantz. 1994. "Some key features of Distributed Morphology." *MIT Working Papers in Linguistics 21: Papers on phonology and morphology*: 275–288.

Harley, Heidi. 1994. "Hug a tree: Deriving the morphosyntactic feature hierarchy." *MIT Working Papers in Linguistics 21: Papers on phonology and morphology*: 289–320.

Harley, Heidi, and Ritter Elizabeth. 2002. "Person and number in pronouns: a feature-geometric analysis" *Language* 78: 482–526. doi:10.1353/lan.2002.0158

Harris, James. 1994. "The syntax-phonology mapping in Catalan and Spanish clitics." *MIT Working Papers in Linguistics 21: Papers on phonology and morphology*: 321–353.

Hayes, Bruce. 1989. "Compensatory Lenghtening in Moraic Phonology." *Linguistic Inquiry* 20: 253–306.

Holmberg, Anders, and Ian Roberts. 2010. "Introduction: parameters in minimalist theory." In *Null Subjects: the structure of parametric variation*, ed. by Theresa Biberauer, Anders Holmberg, Ian Roberts, and Michelle Sheehan, 1–57. Cambridge: Cambridge University Press.

Hubert, Haider, and Rositta Rindler-Schjerve. 1987. "The parameter of auxiliary selection. Italian-German contrasts." *Linguistics* 25: 1029–1055.

Ledgeway, Adam. 2000. *A Comparative Syntax of the Dialects of Southern Italy: A Minimalist Approach*. Oxford: Blackwell.

Legendre, Gerardine. 1989. "Unaccusativity in French". *Lingua* 79: 95–164. doi:10.1016/0024-3841(89)90067-3

Legendre, Gerardine. 2010. "A formal typology of person-based auxiliary selection in Italo-Romance." In *Syntactic variation. The Dialects of Italy*, ed. by Roberta D'Alessandro, Adam Ledgeway, and Ian Roberts, 186–200. Cambridge: Cambridge University Press.

Loporcaro, Michele. 1988. "History and geography of raddoppiamento fonosintattico: remarks on the evolution of a phonological rule." In *Certamen phonologicum. Papers from the 1987 Cortona phonology meeting*, ed. by Pier M. Bertinetto, and Michele Loporcaro, 341–387. Torino: Rosenberg & Sellier.

Loporcaro, Michele. 1997a. "Lenghtening and raddoppiamento fono sintattico." In *The dialects of Italy*, ed. by Martin Maiden, and Mair Parry, 41–51. York: Routledge. Basel-Tübingen: Francke.

Loporcaro, Michele. 1997b. *L'origine del raddoppiamento fono sintattico. Saggio di fonologia diacronica romanza*.

Loporcaro, Michele. 1998. *Sintassi comparata dell'accordo participiale romanzo*. Torino: Rosenberg & Sellier.

Loporcaro, Michele. 2010. "The logic of Romance past participle agreement." In *Syntactic variation. The Dialects of Italy*, ed. by Roberta D'Alessandro, Adam Ledgeway, and Ian Roberts, 225–243. Cambridge: Cambridge University Press.

Manzini, Maria R., and Leonardo M. Savoia. 2005. *I dialetti italiani e romanci. Morfosintassi generativa*. Alessandria: Edizioni dell'Orso.

Nespor, Marina. 1993. *Fonologia*. Bologna: Il Mulino.

Noyer, Rolf. 1997. *Features, Positions and Affixes in Autonomous Morphological Structure*. New York: Garland Publishing.

Passino, Diana. 2012. "The weight of empty structure: Raddoppiamento Sintattico blocking in Teraman Abruzzese". In *Studi e Saggi Linguistici*, ed. by Romano Lazzeroni, and Giovanna Marotta, 105–133. Pisa: ETS.

Pellegrini, Giovanni B. 1977. *La Carta dei dialetti d'Italia*. Pisa: Pacini.

Perlmutter, David M. 1978. "Impersonal Passives and the Unaccusative Hypothesis." *Proceedings of the 4th Annual Meeting of the Berkley Linguistic Society*: 157–189. doi:10.3765/bls.v4i0.2198

Ritter, Elizabeth, and Martina Wiltschko. 2009. "Varieties if INFL: TENSE, LOCATION and PERSON." In *Alternatives to Cartography*, ed. by Jeroen van Craenenbroek, 153–201. Berlin: Mouton de Gruyer. doi:10.1515/9783110217124.153

Ritter, Elizabeth, and Martina Wiltschko. 2010. "The compositon of INFL: An exploration of tense, tenseless languages and tenseless constructions." *LingBuzz/001078*, July 2010.

Schuchardt, Hugo. 1874. "De quelques modifications de la consonne initiale dans les dialectes de la Sardaigne, du Centre et du Sud de l'Italie." *Romania* 3: 1–30.

Sorace, Antonella. 2000. "Gradients in auxiliary selection with intransitive verbs." *Language* 76: 859–890. doi:10.2307/417202

Torcolacci, Giuseppe. 2014. "Il Raddoppiamento Fonosintattico e la codifica di tratti morfosintattici. Il caso dei dialetti italiani meridionali." *L'Italia dialettale LXXV*: 247–271.

Torcolacci, Giuseppe. 2015. *Marking the Default. Auxiliary selection in Southern Italian dialects*. Utrecht: LOT.

Van Valin, Robert D. 1990. "Semantic parameters of split intransitivity." *Language* 66: 221–260. doi:10.2307/414886

Vincent, Nigel. 1988. "Non-linear phonology in diachronic perspective: stress and word-structure in Latin and Italian." In *Certamen phonologicum. Papers from the 1987 Cortona phonology meeting*, ed. by Pier M. Bertinetto, and Michele Loporcaro, 421–432. Torino: Rosenberg & Sellier.

Waltereit, Richard. 2004. "Fossils of a former phonological rule: irregular raddoppiamento fonosintattico." *Philologie im Netz* 29: 40–59.

CHAPTER 7

Partial subject paradigms and feature geometry in Northern Occitan dialects

Michèle Oliviéri, Jean-Pierre Lai and David Heap
Université Côte d'Azur / Université Grenoble Alpes /
University of Western Ontario

In between the minority of Romance languages that have full paradigms of subject clitics (e.g. Standard French) and the unmarked null subject grammars (e.g. Spanish, Italian and most Occitan dialects), a continuum of transitional varieties shows between one and five nominative clitics. Unlike the better documented Northern Italian Dialects, Northern Occitan Dialects have nominative clitics paradigms which typically begin with distinct meteorological subjects. The shape and sequence of the partial nominative clitic paradigms maps the progressive diachronic introduction of contrasts using underspecified monovalent features organized hierarchically to reflect implicational dependencies, following Harley & Ritter's (2002) Feature Geometry. Meteorological subject pronouns play a crucial role in this diachronic progression precisely because they lack morphological features and therefore can map onto non-referential subjects.

Keywords: clitic subjects, null subjects, meteorological verbs, partial paradigms, Occitan

1. Introduction

Within the long-standing debate regarding subject clitics and the so-called Null Subject Parameter, we focus on the emergence of these elements in Romance languages, trying to establish if it can be stated in a general Romance fashion, notwithstanding the great variation observed between dialects.

Considering first the Null Subject Parameter, it is obvious that the change between the two types of grammars is not sudden. Instead we find a gradual transition, displaying many 'mixed' systems, or partial subject paradigms (also known as 'partial Null Subject' systems). Like a number of other scholars (Renzi & Vanelli 1983; Cabredo Hofherr 2004; Manzini & Savoia 2005, among others), we have been

DOI 10.1075/rllt.11.07oli

studying this progression for some time,[1] and yet the remaining questions include: how does this transition work? Is it universal or does it depend on other aspects of grammar? Secondly, as we assume the postulate that diachronic change is revealed by diatopic variation (cf. Gilliéron & Mongin 1905, and more recently Dalbera 2006 or De Vogelaer & Seiler 2012), we claim that, in a homogeneous area, the diversity of the systems illustrates the different stages of a language's evolution.

In this perspective, we examine here the behaviour of Northern Occitan Dialects (NODs) which display partial subject paradigms, contrary to other Occitan dialects which have Null Subject grammars, in order to reconstruct this evolution. As Meyerhoff (1997: 97) underlines: "The place at which systems split also invites attention. Are the restrictions on which subjects may be phonetically null predictable cross-linguistically, or are the constraints always going to be idiosyncratic to each language?"

Indeed, in Northern Occitania, contrary to what has been observed in Northern Italian Dialects (NIDs),[2] the order of appearance of these subject clitics is not the same in all dialects where this transition takes place. Therefore, this question of "the place at which systems split" is the focus of this research.

2. Data from Northern Occitan Dialects (NODs)

The area which interests us in the current study is along the northern boundary of Occitania, which occupies most of Southern France, except the Catalan, Basque, and Ligurian areas, with the addition of a few points in Spain and in Italy, as shown on Map 1.

Map 1 also reveals the situation of Occitan Dialects with regard to Null Subject, as the major part of the area (in white on the map) displays Null Subject grammars, while northern dialects (black area on the map) have full paradigms of subject clitics. Southern Occitan dialects are in this respect comparable to Spanish and Italian, while northern Gallo-Romance dialects resemble French (more or less) in this respect. Table 1 gives some examples of verbal paradigms which are representative of the whole Southern Occitan domain in white on the map.

1. Heap (2000, 2002), Oliviéri (2010, 2011), Heap and Oliviéri (2013).

2. Following Italian scholars, we adopt the acronym NIDs for convenience, in referring to what might more properly be termed a group of related languages.

Map 1. Occitania and the Null Subject Parameter

Table 1. Southern Occitan Dialects: some verbal paradigms (*to be*) from ALF

Dpt	Lectoure	Les Matelles	Villelaure	Les Ternes
	(Gers)	(Hérault)	(Vaucluse)	(Cantal)
P1	suj	suj	sjɛw	ʃu
P2	ɛs	sjes	sjɛs	ʃɔʃ
P3	ɛs	ɛz	ɛz	ej
P4	ɛ̃n	sjẽ	sjɛ̃	ʃẽ
P5	ɛts	sjɛz	sjas	ʃɛʃ
P6	sũ^{n}	su	sũ	ʃũ
it rains	plaɔ	plɔw	plɔw	plɛw

However, as we see on Map 1, there is a grey area in northern Occitania (the focus of Map 2, below) where pronominal subjects begin to appear but do not reach all grammatical persons. This is therefore the zone of interest for us in this study and Table 2 shows some of these NODs paradigms.

Table 2. Northern Occitan Dialects: some verbal paradigms (*to be*) from atlases[3]

	Le Mont-Dore	Eymoutiers	Coussac-Bonneval	Tayac	Vélines	St-Pardoux-La-Rivière
	(1705)	(1604)	(1608)	(338)	(1634)	(1612)
Dpt	Cantal	Haute-Vienne	Haute-Vienne	Gironde	Dordogne	Dordogne
P1	se	jo se	se	sɛj	sej	sø
P2	t se	te se	ty se	tœ se	te ʃɛʲ	ty se
P3	e	ej	u e	ew ej	ej	u / l e
P4	sã	nu sũ	nu sũ	sɔ̃ŋ	sɔ̃	nu sũ
P5	sɛ	vu se	vu se	vuzaw se	bu ʃej	vu se
P6	sɔ̃	sũ	sũ	zi sɔ̃ŋ	sɔ̃	i / la sũ

The available data come from two main sources. First, the atlases (ALF, ALAL, ALLOc, ALG) provide some information but, since their aim was the lexicon and not morpho-syntax, the data remain fragmentary.[4] For instance, ALAL does not have any map with P4 and P5 verbal forms However, although it is not possible to reconstitute the whole paradigm of a verbal conjugation, these data still show some interesting facts (see Table 3). The systems presented here are at the very beginning of the use of subject clitics[5] and the brackets indicate that they are still optional, but we see that (unlike in the NIDs) Person 3 and Person 1 clitics can appear while Person 2 clitics are still absent.

The other source of data is our recent fieldwork[6] in three *Départements* of this region: Corrèze in 2010, Creuse in 2011 and Dordogne in 2013. Most of the data in this paper are from these fieldwork sessions.

In Corrèze, as we will see below, there is no systematic referential subject clitic, but it appears that some of them are just beginning to emerge. It is the case for

3. The localities' numbers are those of the database THESOC (See References): the "THESAURUS OCCITAN" or "THESOC" is a multimedia database dedicated to Occitan Dialects. For a full presentation, see Dalbera et al. (2012).

4. This is the reason why the only full paradigm available is the verb *to be*.

5. It is well known that the emergence of subject clitics is related to the loss of rich verbal morphology (even if there are some exceptions), but the aim of this paper is not to discuss that point, nor which one of both phenomena triggers the other one.

6. Thanks to the French-German project DADDIPRO "Dialectal, acquisitional, and diachronic data and investigations on subject pronouns in Gallo-Romance" (2012–2015), UNS – UMR 7320 (France) & Universität Konstanz (Germany): http://www.agence-nationale-recherche.fr/projet-anr/?tx_lwmsuivibilan_pi2[CODE]=ANR-11-FRAL-0007.

example in Veyrières (loc. 846) or La Chassagne (loc. 849) where a P1 clitic occasionally occurs: [ju], [jø].

Table 3. Northern Occitan Dialects: emergence of subject clitics

	Saint-Victor-La-Riviere (590)	La Sauvetat (588)	Gioux (605)
Dpt	**Puy-de-Dôme**	**Puy-de-Dôme**	**Creuse**
P1			(jo)
P2			
P3 masc	(e)	(ɛ)	(u)
P3 fem	(e)	(ɛ)	(la)
P6 masc		(ɛ)	(o)
P6 fem			

This fact and those in Table 3 reveal a particular pattern in the NODs, different from that of the better-documented NIDs. Crucially the sequence of grammatical persons is not the same as in the NIDs (and in Franco-provençal, see Diémoz 2007), where Person 2 regularly appears first, as shown in Table 4:

Table 4. Northern Italian Dialects: some subject clitics paradigms. Data from Manzini and Savoia (2005) and THESOC

	Breil	Airole	Rochetta Nervina	Tende	Saorge
	to leave	*to sleep*	*to sleep*	*to sing*	*to understand*
P1	ˈpaʀti	ˈdɔrmu	ˈdɔrmu	ˈkantu	e kaˈpiʃə
P2	ti ˈpaʀte	ti ˈdɔrmi	ti ˈdɔrmi	ti ˈkanta	ti kaˈpiʃə
P3	aʀ ˈpaart	a ˈdɔrme	u / a ˈdɔrme	aɻ / a ˈkanta	ə / a kaˈpiʃə
P4	paʀˈtimª	durˈmemu	dorˈmemu	kanˈtamu	e kapiˈʃemə
P5	paʀˈti	durˈmei	dorˈmei	kanˈtai	e kapiˈʃei
P6	ˈpaʀtu	in ˈdɔrme	i ˈdɔrme	li / le ˈkantaᵑ	e kaˈpiʃuŋ
it is raining	ˈtʃɔu	ˈtʃø:ve	ˈtʃø:ve	aɻ ˈtʃou	ə ˈtʃɔu

Notice that in the NIDs, meteorological subjects appear late, with or after other P3 subjects, and when they appear, they are homophonous with P3 masculine pronoun, as in French.

3. Previous analyses

Based primarily on the behaviour of subject pronouns across NIDs, several different progressions have been previously modelled, in terms of the order of appearance of

subject pronouns, typically beginning with P2. Table 5 summarizes these analyses, noting a division into two blocks (grey vs. white), as pointed out by Heap (2000) among others. The first block (P2, P3, P6) is most likely to appear first in split subject pronoun paradigms, while the second block (P4, P1, P5) appears last:

Table 5. Previous analyses of the progressive emergence of subject clitics

	Renzi & Vanelli	Heap	Cabredo Hofherr	Oliviéri
	(1983)	(2000)	(2004)	(2009)
stage 0				ø
stage 1	P2	P2, P3, P6	P2	P2
stage 2	P3		P3	P3
stage 3	P6		P6	P6
stage 4	P5	P4, P1, P5		Expletive
stage 5	P4		P1	P1
stage 6	P1		P4 and P5	P4 and P5

Heap (2002) notes the split between Block A pronouns (2sg, 3sg & 3pl) which are favoured, and Block B pronouns (1pl, 1sg & 2pl) which are disfavoured, as well as the tendency for 2sg must be present if there are two or more subject clitics.

But these generalizations, which imply that P2 subject clitic is among the first SCl to appear but also that P1 is in the second block, cannot stand since we have seen that the NODs behave differently. Taking into account all the NODs (and NIDs) configurations, Oliviéri (2011: 248) models the emergence of subject clitics along with the well-known loss of verbal agreement morphology in Romance, and shows that what is relevant is not the presence of the clitic, but rather the functional need to distinguish grammatical persons. Oliviéri goes on to describe these emerging subject pronouns using a matrix of binary features (as in Table 6).

Under such an analysis, it is a mere coincidence that the progression proposed, summarized in (1),[7] involves the successive introduction of the features in the system.

(1) [Person] > [Speaker] > [Number] and [Gender]

And the expletive can appear freely as soon as P3 clitic is present in the system (since they are homophonous with the P3 masculine).

This progression adequately describes the range of attested NID paradigms, for the acquisition of French,[8] and for the most part NOD paradigms as well – ex-

7. See Oliviéri (2011: 249–250) for a detailed presentation.

8. See Palasis (2010) and the results of the DADDIPRO project (Oliviéri et al. to appear).

Table 6. Binary features

	[± Person]	[± Speaker]	[± Feminine]	[± Plural]
P1	+	+	0	−
P2	+	−	0	−
P3 masc	−	0	−	−
P3 fem	−	0	+	−
P4	+	+	0	+
P5	+	−	0	+
P6 masc	−	0	−	+
P6 fem	−	0	+	+
Expletive	0	0	0	0

cept, as we will see, for the expletive. However, this configuration is insufficiently constrained, as it significantly overpredicts the number of possible paradigms vs. those that are actually attested.

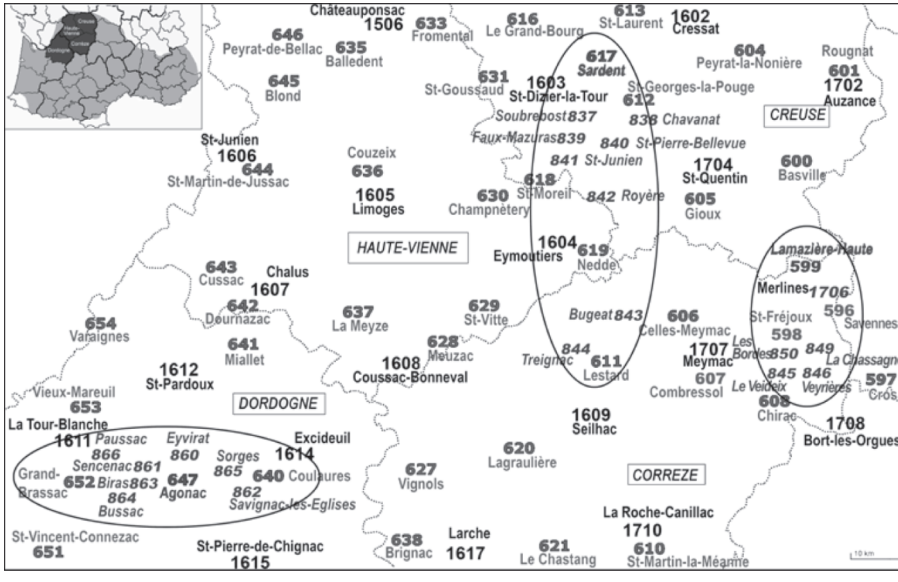
4. *ko*: A different subject pronoun[9]

Map 1 shows a grey area, the transitional zone with partial Null Subject systems, where the phenomena studied here are prevalent, particularly in the *Départements* of Corrèze, Creuse, Haute-Vienne and Dordogne. Map 2 shows the network of atlas data points combined with our new complementary survey areas (indicated by the three ovals).

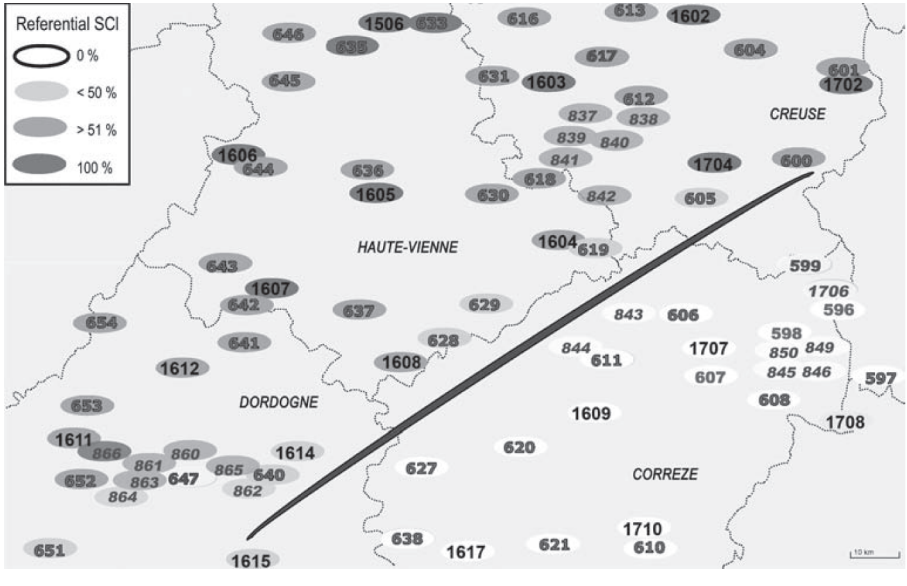
Map 3 shows the frequency of referential subject clitics: the further north we go, the darker the points are i.e. the higher the frequency of subject clitics. These data allow us to draw an isogloss dividing the Null Subject area from the area with subject pronouns.

It is however striking that in this area, meteorological subjects are the first subject pronouns to appear (cf. also Kaiser et al. 2013). As we see in these Corrèze paradigms in Table 7, even when there are no referential subject clitics, with verbs like 'to rain' or 'to hail', a special pronominal element appears: the pronoun *ko*.

9. There are of course a wide variety of meteorological predications in the typology of impersonals (see Mettouchi & Tosco 2011), but here we are only concerned with meteorological subject clitics in these Romance varieties.



Map 2. NODs network



Map 3. Referential subject clitics in these NODs

Table 7. 'To go' vs. meteorological verbs in Corrèze[10]

	Lamazière-Hte	**Bugeat**	**Treignac**
	(599)	**(843)**	**(844)**
P1	vˈɔu	vˈau	vˈɔu
P2	vˈa	vˈa	vˈa
P3	vˈaj	vˈa	vˈaj
P4	anˈɛ̃	anˈã	anˈã
P5	anˈɛ	anˈa	anˈa
P6	vˈɔ̃	vˈɔ̃	vˈɔ̃
it is raining	kɒ plˈø	kɔ plˈø	ka plˈøw
it is hailing	kɒ gʀˈɛlo	kɔ gʀˈɛlo	ka gʀˈɛlo

This fact represents a major difference between NODs and NIDs (where meteorological verbs have either no expletive subject, or an expletive subject which is identical to the P3 masculine form).

Nevertheless, this element *ko* does not always appear and Table 8 illustrates its frequency in our interview survey data, according to the different contexts. The first column lists the sentences of our questionnaire, showing that *ko* is more frequent in root sentences and in the present tense (and it seems that there is no difference with coordinated and circumstantial contexts). However, it is less frequent when the sentence is negative or in another tense, and even less in embedded contexts. Moreover, we see diatopic variation between the three *Départements*, with categorical *ko* usage generally increasing from Corrèze to Creuse, and Dordogne in the middle.

Turning to other meteorological expressions (with the verb *faire* 'to do'), we observe (Table 9) that *ko* is less frequent, but that the phenomenon spreads gradually to other constructions, with the same proportional relationship between the three geographical areas.

Some examples are given in Table 10 where the first line shows *ko* usage with a meteorological verb and the following lines illustrate *ko* with weather expressions (of the type *il fait + froid, nuit, beau…*). The subject *ko* is always present with *to rain*, and variable with the others and we observe the same progression as before, from Creuse to Dordogne to Corrèze.

Comparing Maps 4 and 5, it appears that the phenomenon starts with meteorological verbs and then it extends to other meteorological expressions.

10. Although the verb *to go* is an irregular one, it is representative of the global behaviour of the verbal morphology in these dialects, i.e. rich verbal morphology and no subject clitics.

Table 8. Frequency of *ko* with weather verbs

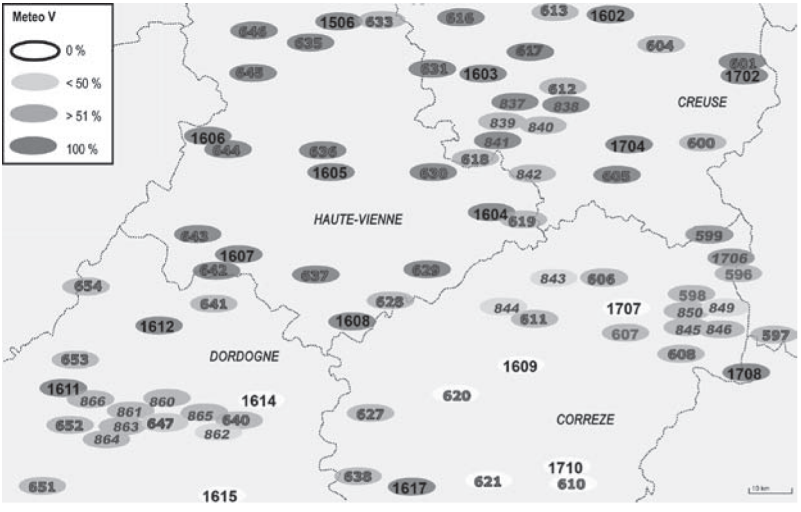
	Corrèze	Dordogne	Creuse	Type	Context	Mood-Tense
Oui, ça pleut bien.	**100%**	**100%**	**100%**	root		Ind Prst
Il pleut?	**100%**	**100%**	**93%**	root	interr	Ind Prst
Il pleut.	**88%**	**100%**	**100%**	root		Ind Prst
Moi ça me plaît pas quand il grêle.	**100%**	75%	**86%**	circ		Ind Prst
Il fait froid et pleut.	**81%**	75%	**100%**	coord		Ind Prst
Il ne pleut plus maintenant.	**57%**	**63%**	**100%**	root	neg	Ind Prst
Il a beaucoup plu hier.	50%	**56%**	**100%**	root		pc (+aux)
Avant il neigeait tous les ans.	29%	**75%**	**86%**	root		Impft
Non, il ne pleut pas.	50%	50%	**86%**	root	neg	Ind Prst
J'aimerais qu'il pleuve!	0%	50%	**64%**	embedded		Subj Impft
Il faut qu'il pleuve!!!	0%	25%	**79%**	embedded		Subj Prst
Il faudrait qu'il ne pleuve plus.	14%	13%	**71%**	embedded		Subj Impft
Espérons qu'il ne pleuve pas!	0%	25%	**64%**	embedded		Subj Prst
Tu crois qu'il pleuvra demain?	0%	13%	50%	embedded		Fut

Table 9. Frequency of *ko* with other meteorological expressions

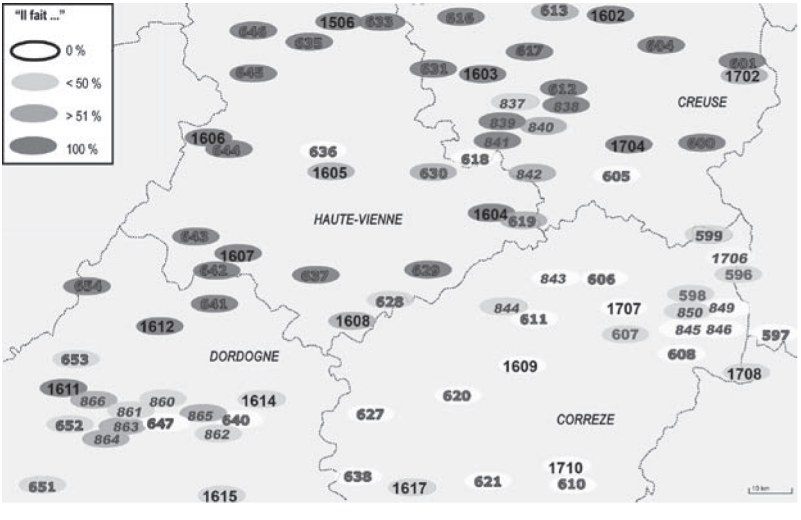
	Corrèze	Dordogne	Creuse	Type	Context	Mood-Tense
Il fait froid.	25%	**75%**	**86%**	root		Ind Prst
Il fait nuit.	11%	**63%**	**83%**	root		Ind Prst
Quel temps il fait?	13%	25%	**86%**	root	interr	Ind Prst
S'il fait beau demain,…	0%	25%	50%	circ		Ind Prst

Table 10. *ko* with *faire*

	Veyrieres	Lamaziere-Haute	Sorges	Biras	Sardent	Faux-Mazurat
	(846)	(599)	(865)	(836)	(617)	(869)
Dpt	**Corrèze**	**Corrèze**	**Dordogne**	**Dordogne**	**Creuse**	**Creuse**
Il pleut	**kɔ** plø	**ka** plø	**ko** plo	**ko** plow	**ka** plo	**kɒ** plo:
Il fait froid.	faj fʁɛ	**ka** fɛ fʁɛ	**ko** faj fʁɛ	**ko** faj frɛ	**ka** fe frɛ	**kɒ** fe fʁɛ
Il fait nuit.	faj njøj	fa ni:	**ko** faj nɛ	**ko** faj ne	**ka** fe nø:j	**kɒ** faj nɛ
Quel temps il fait?	ka tɛ̃ fɛ	kɔw tɛ̃ fɛ	kaw tɑ̃ faj **ko**	kaw tɛ̃ faj	kaw tɛ̃ **k** fɛ	kaw tɛ̃ faj **kɒ** kaw tɛ̃ **kɒ** fɛj
S'il fait beau demain,…	ʃi fɛ be tɛ̃	ʃi faj bu	si faj bʁave tɛ̃	ʃi **ko** faj bũ	si **ka** fɛ bu	si **kɒ** fe bu



Map 4. *ko* with meteorological verbs



Map 5. *ko* in meterological expressions with *faire*

With other impersonal constructions (Table 11), we see the same progression with *il est*, while with *sembler* and *paraître* (to seem), *ko* is just starting to show up.

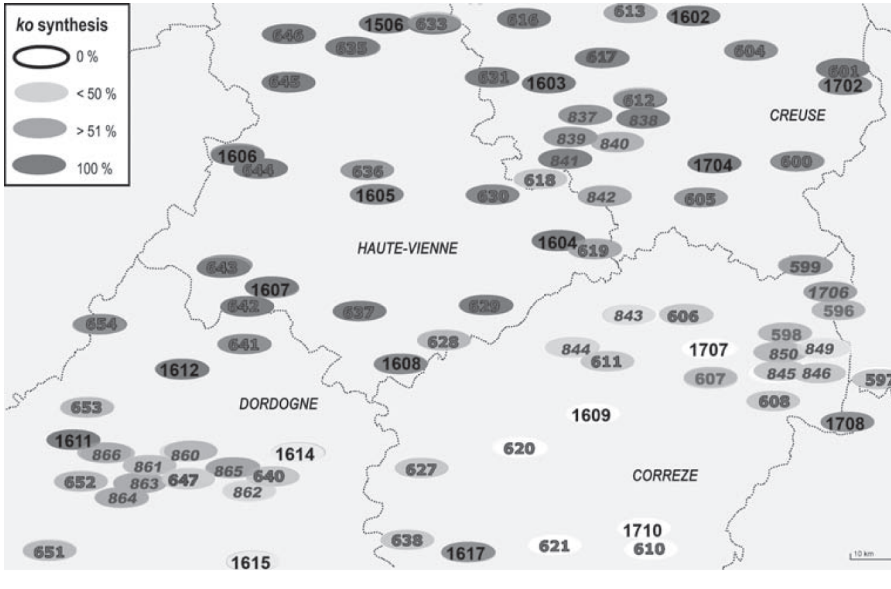
Once again we observe the same tendencies from a diatopic point of view and (2) illustrates the different stages of the diachronic emergence of *ko*.

Table 11. *ko* with impersonal verbs

	Bugeat	Lamazière-Haute	Paussac et St Vivien	Sorges	Royère-de-Vassivière	
	(843)	(599)	(866)	(865)	(842)	
Dpt	**Corrèze**	**Corrèze**	**Dordogne**	**Dordogne**	**Creuse**	
Il est tard.	ɛ taʁ	ɛ taːʁə	**k** ɛj tɐr	**k** ei tɐʁ	**k** ɛ tæʁə	*"it is late"*
Il paraît…	paʁɛ	**ka** paʁɛ	pɐre	**ko** pɐʁe	paʁe	*"it seems…"*
Il me semble…	mə ʃɑ̃bla	mə sɑ̃blə	me ʃɛ̃ble	mə ʃɛ̃blə	**ka** mø sɛ̃blə	*"it seems to me…"*

(2) Stage 1: Meteorological Verbs
Stage 2: Meteorological expressions
Stage 3: Impersonal Verbs 1 *(it is late)*
Stage 4: Impersonal Verbs 2 *(it seems)*

Map 6, which synthesizes the different occurrences of *ko*, also shows the same trend: Corrèze > Dordogne > Creuse. The comparison between the emergence of referential subject clitics (Map 3) and that of *ko* shows a highly significant contrast between the two, and the isogloss of the Map 3 would not correspond to any clear dividing line on Map 6.



Map 6. Increasing presence of *ko*

However, on both maps we see that Corrèze shows more archaic grammars, followed by the ones in Dordogne at a later stage, then the ones in Creuse, where the grammars are at an even more innovative stage relative to the others. These transitional dialects lead us to question whether the sequence of introduction of subject clitics established earlier on the basis of NIDs is applicable here or needs to be modified.

5. Beyond binary features: A proposed feature geometry

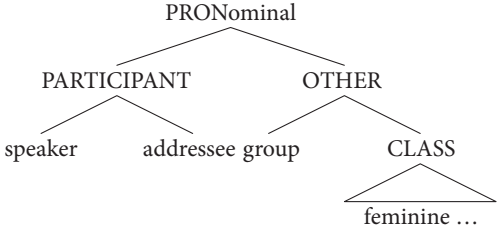
Recalling earlier analyses of pronoun paradigms in terms of binary features (Table 6 above), we note that this approach leaves many co-occurrence restrictions and systematic redundancies unexplained. For example, the feature [+ Fem] only applies in the context of [– Pers], and is always unspecified [0 Fem] for the [+ Pers] forms, i.e. first and second persons. Similarly, the feature [+ Spkr] is never specified for [– Pers] i.e. third person forms, and only specified for [+ Pers] forms. But can the fact that [+ Spkr] and [+ Fem] are in complementary distribution across Romance really be merely coincidental?

Furthermore, if we treat this set of binary features as freely combinable, they have the effect of overpredicting the possible subject pronoun paradigms. There is nothing in principle that prevents combinations like [+ Pers] [+ Spkr] [+ Fem] or [0 Pers] [0 Spkr] [+ Fem] [+ Plur], even though such pronouns seem unattested. The combinatorial potential of four binary features is underutilized by the attested forms: freely combined 2^4 features should give us 16 possible forms (or even more, if the value [0 feature] is also allowed to contrast freely with other combination), whereas the attested paradigms for the systems under study never have more than eight contrasting forms (or nine, if an expletive is included). But many of these combinations never seem to occur, and indeed seem logically unlikely. So, how can we prevent the grammar from predicting such combinations?

A more constrained theory of features should allow us to predict more closely the range of attested paradigms, and also their development, by building dependency relationships directly into the representation of features. A binary feature matrix cannot express this sort of restriction without resorting to stipulations. In order to encode such redundancy relationships between features, we need a geometric approach where structural dependencies are directly reflected in the internal hierarchical structure of pronominal elements, as in (5), below. Note that in a morphological feature geometry like the one proposed by Harley & Ritter (2002), the use of monovalent (rather than binary) features predicts a more constrained set of grammars.

We propose a model of hierarchical organization for morphological features, inspired by Harley & Ritter (2002), but simplified and adapted for our purposes as in (3). We opt for this model which encodes entailment relationships and allows underspecification, but unlike Béjar (2003) we do not posit an intermediate category node π between the root node R and PARTICIPANT. We also propose that the root node in the feature geometry should not be Referring Expression (because some pronouns, in particular meteorological subjects, are not referential) but rather just PRONominal. We propose that the sister node to PARTICIPANT should be simply OTHER (since third persons i.e. non-persons or OTHERS may or may not be individuated, while PARTICIPANTS on the other hand are inherently individuated).

(3) Feature geometry



Heap & Oliviéri (2013) examine cases of NODs where the first clitic to appear can be either a PARTICIPANT (P1 speaker or P2 addressee) or a P3 (OTHER). But what happens when the first SCL is *ko*?

The origin of *ko* is the general Occitan demonstrative *acò* [aˈko] "that" (Ronjat 1937: 86–92), which is normally stressed and deictic. Where other Occitan dialects have the usual deictic [aˈko], the NODs show *ko*.[11] Indeed, in the area studied here, *ko* also appears with this meaning as an object thus in stressed position, alternating with a reinforced form [koˈki] or [ko d aˈki]:

(4) n ublidaˈʁɛ̃ⁿ dzamaĭ ˈko — *Lamazière Haute* (599)
nu n ublidøˈʁɔ̃ ʒaˈme køˈki — *Saint-Pierre-Bellevue* (840)
n ɔwblidəˈrɔ̃ zəˈmɛj ko d ɑˈki — *Biras* (863)
'We will never forget that.'

11. It is still not clear whether this is due to a general mechanism of apheresis in this area (where the verbs [aˈna] 'to go' or [ariˈba] 'to arrive' are respectively [ˈna] and [riˈba], see also Ronjat (1937: 90)) or to a new distinction between a stressed [aˈko] and an unstressed [ko] (as it seems to be the case further north).

The phonological/prosodic reduction to weak *ko* seems to correspond to a loss of morphological (and semantic) specification, but what exactly is the feature that distinguishes (strong) [(a)ˈko] from (weak) [ko]? How does the loss of this featural specification allow *ko* to become specialized as a non-referential subject e.g. for meteorological verbs?

Since [(a)ˈko] is in principle deictic (and referential), let's assume then that if it loses some properties related to deixis, we are left with a pronominal element which doesn't "point" at anything and therefore doesn't refer to anything: a non-referential element, like the impersonal subject of meteorological verbs. In many grammars, such a non-referential element would be a null subject par excellence, but if we are going to put a pronominal element in this position, we need to be sure that is doesn't refer, i.e. has no phi-features. According to Burzio (1992), reflexives have in common with (some) impersonals their featural make-up: they have no phi-features i.e. they would be just PRONominal, the bare root node in this particular feature geometry.

In order to distinguish these two maximally underspecified pronominals, we have to suppose that some feature distinguishes reflexive clitics (like *se*) from impersonal clitics like *ko*. While underspecified clitics like *se* exist throughout Romance languages with a range of functions (Bruhn de Garavito, Lamarche & Heap 2002), nominative impersonals of the *ko* type, which are not homophonic with P3 masculine singulars, seem relatively rare and peculiar to a few restricted areas like the one under study. Since we need to keep *se* in the overall paradigm of clitics throughout the area, this suggests we need to add a specification to this new *ko* clitic that is not present in the make-up of SE.

The one thing that the new *ko* can be that *se* cannot be is a subject, so an intuitive feature to introduce here would be something marking a pro-nominal which is NOMinative, which can in turn serve as the root node for the whole (new) series of nominative clitics, such that the root node in (3) would be NOM (rather than PRON).[12]

6. Building up contrast in paradigms

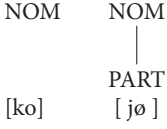
When *ko* is the only clitic, it contrasts with all PART pronouns, because it lacks PART, i.e. cannot be either P1 or P2, and with all P3 pronouns, because it lacks OTHER, and cannot refer:

12. The question of whether accusative and dative clitic paradigms are distinguished by means of different root nodes or by privative features lower down in the geometry is left for future research.

(5) 1SCL system NOM
[ko]

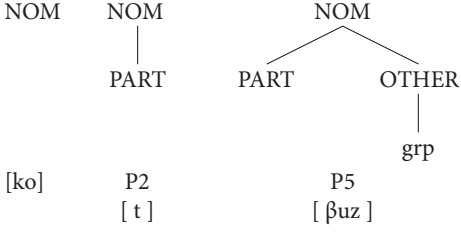
A paradigm with two subject clitics introduces a single contrast: *ko* vs. participant, with P1 representing PART, a system represented by the paradigm used at Veyrières (846):

(6) 2 SCL system



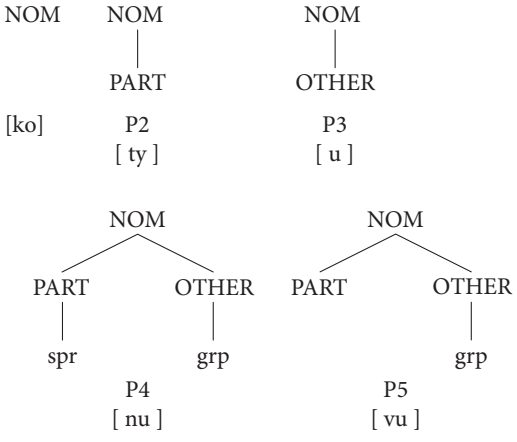
From the point of view of contrast, second clitic could as easily been the other PART person, P2, and indeed this is what we find in the transition to a three subject clitic paradigm, as at Saint-André-de-Cubzac (417), where *ko* is joined by one PART clitic (P2) and its plural counterpart (P5). The crucial thing is the step-wise building up of contrast, one feature at a time:

(7) 3 SCL system



To date, we have found no four-clitic paradigm with *ko*. This appears to constitute an empirical gap in the range of attested grammars. There are however a couple of possible five clitic systems attested, both of which build on the direction established by the three clitic system in (7). At Coussac-Bonneval (1608), we find an additional contrast in the singular (P3 vs. P2) and an additional one in the plural (P4 vs. P5), as in (8):

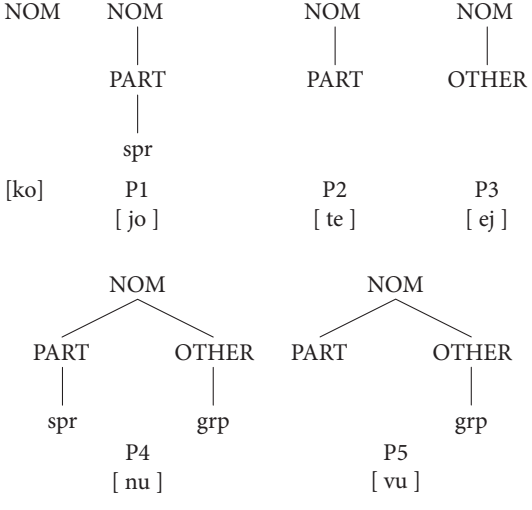
(8) 5 SCL system



The other five-clitic system at Tayac (338) is not shown here, but it is very similar, the only difference being that the additional contrast is between P5 and P6.

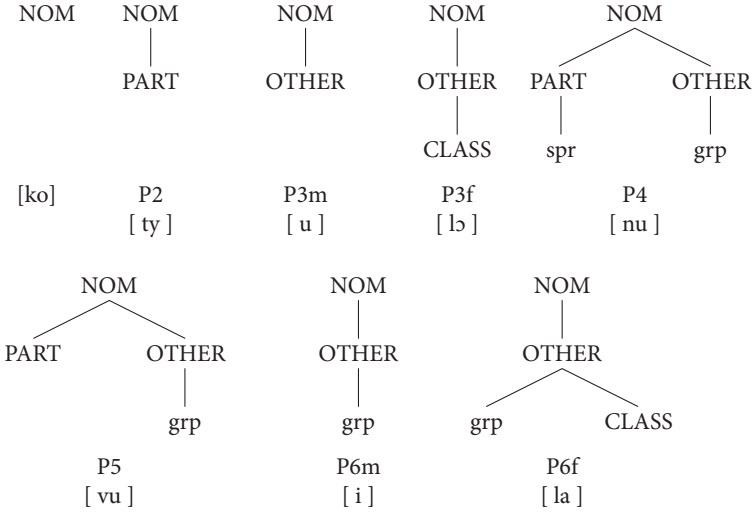
The addition of one more contrast brings us to a six clitic system, with *ko* plus all persons except P6, are found at Eymoutiers (1604):

(9) 6 SCL system



We once again find a gap in the continuum, with no seven clitic system extant in our data so far. We do find however an eight clitic system, with *ko* and all persons, including the addition of a masculine-feminine contrast in the singular and the plural, but not P1, at Saint-Pardoux-la-Rivière (1612):

(10) 8 SCL system



Finally, the full paradigm of nine clitics, with all grammatical persons plus a gender distinction in P3 and P6 as well as *ko*, is found at Les Chiers-Sardent (617), not represented here.

While there are a couple of apparent 'gaps' in the range of attested partial paradigms of subject pronoun in the NODs area (possibly corresponding to dialects which have disappeared or which we have no data for), the extant data strongly suggests that the passage for null subject grammars towards a full paradigm of subject clitics follows different pathways when the starting point is a meteorological subject like *ko*.[13] While there is some variability in possible pathways, the overall path is relatively constrained: there are many possible combinations of grammatical persons which never occur.

7. Conclusion

Just as the range of extant diatopic variants seems to correspond to historical stages in the development of subject clitic paradigms, a number of the stages outlined above are strikingly analogous to the findings of Tosco's account (2005, 2007) of the development of Cushitic 'Secondary Subject Markers' (or SSMs) using a similar feature geometry. For Tosco, the first stage involves the "insertion of a bare Part

13. The status of this meteorological subject *ko* could be analysed (as proposed by Kaiser et al. 2013) as a quasi-argument, i.e. not referential, but argumental (following a typology adapted from Chomsky 1981 by Cabredo-Hofherr 2000).

node, either Speaker or Addressee, and Singular […]; this is the minimal SSMs system." (Tosco 2007: 143). In the case of NODs, this first stage is preceded by the appearance of bare NOM in the form of *ko*, but after that the remaining stages follow the same general pattern of gradual complexification (echoing (1) above) as Tosco traces, for an entirely different family of closely related grammars. Of course, regarding diachronic reconstruction, "any 'proof' which can be inferred from Feature Geometry is at best an indirect one, which gets its value only from the lack of direct, historical data" (Tosco 2007: 149).

Since the possible pathways followed in the development of partial subject pronoun paradigms are constrained by something other than chance, it seems plausible that this principle of hierarchical dependency relationships between privative features, as opposed to a free combination of binary features, should be seen as a property of universal grammar. On the other hand, things that vary between grammars, including details of which features are expressed and which may be underspecified in specific grammars, and whether a non-referential meteorological subject is available or not as a starting point, would have to be parameters that can be set for each individual grammar. The default values would be (micro-)parametric choices that vary between grammars in order to determine the precise shape of each paradigm throughout the continuum. While there is different starting point in the sequence, i.e. *ko*, a bare NOM clitic, the end of the process involves the final addition of P1, a contrast which also typically arrives late in the NIDs as well. Further research will show how an underspecified feature geometry can be used to model the different development pathways followed by subject paradigms in NIDs as well as in NODs.

References

ALF = Gilliéron, Jules & Edmond Edmont. *Atlas Linguistique de la France*, Paris, Champion, 1902–1910.

ALAL = Potte, Jean-Claude. *Atlas Linguistique et ethnographique de l'Auvergne et du Limousin*, Paris, CNRS, 1975–1992.

ALLOc = Ravier, Xavier. *Atlas Linguistique et ethnographique du Languedoc Occidental*, Paris, CNRS, 1978–1994.

ALG = Séguy, Jean. *Atlas Linguistique et ethnographique de la Garonne*, Paris, CNRS, 1954–1973.

THESOC = Dalbera, Jean-Philippe, Dominique Strazzabosco, Michèle Oliviéri, Pierre-Aurélien Georges & Guylaine Brun-Trigaud, *Thesaurus Occitan*, <http://thesaurus.unice.fr>, 1992.

Béjar, Susana. 2003. *Phi-Syntax: A Theory of Agreement*. PhD. thesis, University of Toronto.

Bruhn de Garavito, Joyce, Jacques Lamarche & David Heap. 2002. "French and Spanish Se: Underspecified, not Reflexive." *2002 Proceedings of the Canadian Linguistic Association*.

Burzio, Luigi. 1992. "On the morphology of reflexives and impersonals." In *Theoretical analyses in Romance linguistics. Current issues in linguistic theory* 74, ed by Christiane Laeufer & Terrell A. Morgan, 399–414. Amsterdam: Benjamins.

Cabredo Hofherr, Patrizia. 2000. *La passivation des intransitifs en allemand et le statut des explétifs*. Paris: Université Paris VII – Denis Diderot, U.F.R. de linguistique.

Cabredo Hofherr, Patrizia. 2004. "Les clitiques sujets du français et le paramètre du sujet nul." *Langue Française* 141: 99–109. doi: 10.3917/lf.141.0099

Chomsky, Noam. 1981. *Lectures on Government and Binding. The Pisa Lectures*. Second revised edition 1982. Dordrecht: Foris.

Dalbera, Jean-Philippe. 2006. *Des dialectes au langage: Une archéologie du sens*. Paris: Champion.

Dalbera, Jean-Philippe, Michèle Oliviéri, Jean-Claude Ranucci, Guylaine Brun-Trigaud & Pierre-Aurélien Georges. 2012. "La base de données linguistique occitane THESOC. Trésor patrimonial et instrument de recherche scientifique." *Estudis Romànics* 34: 367–387.

De Vogelaer, Gunther & Guido Seiler (eds.). 2012. *The Dialect Laboratory*. Amsterdam: Benjamins. doi: 10.1075/slcs.128

Diémoz, Federica. 2007. *Morphologie et syntaxe des pronoms personnels sujets dans les parlers francoprovençaux de la Vallée d'Aoste*. Tübingen: A. Francke Verlag.

Gilliéron, Jules & Jean Mongin. 1905. *"Scier" dans la Gaule romane du Sud et de l'Est*. Paris: Champion.

Harley, Heidi & Elizabeth Ritter. 2002. "Person and number in pronouns: A feature-geometric analysis." *Language* 78: 482–526. doi: 10.1353/lan.2002.0158

Heap, David. 2000. *La variation grammaticale en géolinguistique: les pronoms sujets en roman central*. München: Lincom Europa.

Heap, David. 2002. "Split subject pronoun paradigms : Feature geometry and underspecification." In *Current Issues in Romance Languages : Selected Papers from the 29th Linguistic Symposium on Romance Languages (LSRL), Ann Arbor, 8–11 April 1999*, ed. by Teresa Satterfield, Christina M. Tortora, and Diana Cresti, 129–144. Amsterdam-Philadelphia: Benjamins. doi: 10.1075/cilt.220.10hea

Heap, David & Michèle Oliviéri. 2013. "On the emergence of nominative clitics in Romance dialects." Paper at the Workshop European Dialect Syntax VII, University of Constance, 13–15 juin 2013.

Kaiser, Georg A., Michèle Oliviéri & Katerina Palasis. 2013. "Impersonal constructions in northern Occitan." In *Current Approaches to Limits and Areas in Dialectology*, ed by Xosé Afonso Álvarez Perez, Ernestina Carrilho & Catarina Magro, 345–366. Cambridge: Cambridge Scholars Press/CSP.

Manzini, M. Rita & Leonardo M. Savoia. 2005. *I dialetti italiani e romanci: morfosintassi generativa*. Alessandria: Edizioni dell'Orso.

Mettouchi, Amina & Mauro Tosco, 2011. "Impersonal configurations and theticity: the case of meteorological predications in Afroasiatic. " In *Impersonal Constructions: A cross-linguistic perspective*, ed. by Andrej Malchukov & Anna Siewierska, 307–322. Amsterdam: Benjamins. doi: 10.1075/slcs.124.11met

Meyerhoff, Miriam. 1997. *'Be I no GAT': constraints on null subjects in Bislama*. Philadelphia: University of Pennsylvania Thesis.

Oliviéri, Michèle. 2010. "From Dialectology to Diachrony: Evidence from Lexical and Morpho-Syntactic Reconstruction in Romance Dialects." *Proceedings of Methods XIII. Papers from The Thirteenth International Conference on Methods in Dialectology, 2008*, ed. by Barry Heselwood and Clive Upton, 42–52. Frankfurt: Peter Lang.

Oliviéri, Michèle. 2011. "Typology or Reconstruction: the Benefits of Dialectology for Diachronic Analysis." In *Romance Languages and Linguistic Theory 2009. Selected papers from 'Going Romance' Nice* 2009, ed. by Janine Berns, Haike Jacobs, and Tobias Scheer, 239–253. Amsterdam: Benjamins. doi: 10.1075/rllt.3.14oli

Oliviéri, Michèle, Georg A. Kaiser, Michael Zimmermann, Katerina Palasis, and Richard Faure. à paraître. 2015."Quand la dialectologie, la diachronie et l'acquisition se parlent : Étude comparative des pronoms sujets en occitan et en français." In *SyMiLa 2015 : La microvariation syntaxique dans les langues romanes de France*, ed. by Patrick Sauzet, and Jean Sibille. Limoges: Lambert Lucas.

Palasis, Katerina. 2010. *Syntaxe générative et acquisition: le sujet dans le développement du système linguistique du jeune enfant*. Villeneuve d'Ascq: ANRT Diffusion.

Renzi, Lorenzo & Laura Vanelli. 1983. "I pronomi soggetto in alcune varietà romanze." In *Scritti linguistici in onore di G.B. Pellegrini*, 42–56. Pisa: Pacini.

Ronjat, Jules. 1937. *Grammaire Istorique des Parlers Provençaux Modernes*. Genève-Marseille: Slatkine / Laffitte Reprints.

Tosco, Mauro. 2005. "La naissance d'une catégorie morphologique: les clitiques sujet entre couchitique et langues romanes." *Faits de langues* 26: 203–215.

Tosco, Mauro. 2007. "Feature-geometry and diachrony." *Diachronica* 24 1: 119–153. doi: 10.1075/dia.24.1.06tos

CHAPTER 8

Automatic detection of syntactic patterns from texts with application to Spanish clitic doubling

Bruno Estigarribia
University of North Carolina Chapel Hill

We developed an automated algorithm to retrieve direct object clitic doubling (DOCLD) examples in Spanish data from texts and the web. We focused on the Rioplatense dialect, where this kind of doubling is rather common. Given an electronic text, our procedure has two steps: first, tagging the text with an available part-of speech (PoS) tagger (TreeTagger), then inputing the tagged text into java-based code that extracts all sentences containing direct object clitics and attempts to match each clitic to a candidate doubled NP in its sentence. Identification of DOCLD cases in a short story (edited text) was 100%, whereas on unedited, raw text it was only 50%. Missing DOCLD cases are mainly caused by misspellings and lack of punctuation in the raw texts. We discuss how to improve accuracy mainly by reducing the number of false negatives.

Keywords: Web-as-corpus, automatic syntactic analysis, Rioplatense Spanish, corpus linguistics, parsing, pattern identification

1. Corpus linguistics and the World Wide Web

The advent of the World Wide Web (WWW) with ever greater data storing and computing capabilities has radically changed the practice of corpus-based and computational linguistics (the "web-as-corpus" approach; see Gries, Wulff & Davies 2009; Gries 2009 for discussion).

This "web-as-corpus" approach has many advantages over traditional corpora. First, web data is orders of magnitude bigger than constructed corpora.[1] Web data is also free, publicly available, and continuously updated (Fletcher 2012; Kilgarriff

1. As of May 1st, 2017, the WWW can be estimated to contain at least 4.49 billion pages (http://www.worldwidewebsize.com/)

DOI 10.1075/rllt.11.08est

& Grefenstette 2003), making possible the examination of language change in great detail in short time periods. The web also represents many dialects of standard varieties and their contact in a virtual community, a kind of data that is unique. Uses of the WWW for corpus linguistics are wide-ranging, and this variety further attests to the usefulness of mining the web for language data.[2] One of the simplest ways in which linguists use the WWW is by querying search engines, e.g. counting Google hits for different word forms (Fletcher 2012). There is a more sophisticated emerging body of work using Twitter (Gimpel et al. 2011),[3] but Twitter is not the only WWW content that is useful for linguistic research and is clearly inadequate for some goals. Twitter posts ("tweets") are limited to 140 characters, and users cannot really fully engage in "conversations". This limits Twitter's usefulness for problems such as basic language documentation and syntactic description, machine translation, and stylistic and discourse studies. For these, Facebook and blogs, for example, represent a better body of data.

For all these advantages, web data are underutilized by syntacticians, compared to other practitioners of linguistics.[4] The greater part of contemporary generative theory is based not on corpus data, but on introspective data. Here again is where web data can greatly enrich linguistic theory (more advantages are made clear in Section 2).

Of course, the web has known issues. A web corpus may not be representative of a language variety (as opposed to a carefully constructed corpus) (Kilgarriff & Grefenstette 2003; Baker 2010). Web language is a little-studied "third variety", not really written, not really spoken, but somewhere "in between". The high percentage of "spamming" pages biases word-frequency and other counts. Finally, and most relevant for this project, there is the problem of web-specific interference: typos, idiosyncratic spellings, and idiosyncratic, variable formatting. As a result of these issues, currently there are no automated natural language processing (NLP) tools or protocols to identify, clean up, and analyze web language data in a particular language. Most data clean-up is made through very time intensive human pre-processing (see the discussion in Fletcher 2012).

---

2. Web data is increasingly used in lexicography, in machine and human translation, in language teaching, for protection of endangered languages and language policy. Studies of ideology, sentiment analysis, genre, style, register, intertextuality, metaphor, and discourse analysis also avail themselves of web data, as do forensic linguistics and marketing research.

3. This is due in large part to Twitter's very liberal terms of service and user-friendly application programming interface (API) (Russell 2013).

4. These two paragraphs should not be construed as either claiming that corpus data should replace introspective data, or that no generative linguists ever use corpus data. Rather, it calls for a stronger synergy between the two.

In this paper, we will describe our efforts to build an algorithm that would, in the ideal case, scour the web for Spanish texts and identify a specific syntactic construction in these texts with a minimum of human intervention. We have chosen as a case study the problem of identifying Direct Object Clitic Doubling (DOCLD) in electronic texts. For this study, we assumed the texts as given, that is, we do not examine the issue of automatizing the process of "scouring" the web and retrieving data from different web sources. We focus instead on the specifically linguistic problems that the identification of DOCLD poses.

2. Our case study: Identifying DOCLD from web texts

DOCLD is a syntactic configuration where the direct object of a transitive verb is encoded both by a DO clitic and by another constituent (the "double") in argument position (in this study we will not differentiate between DOCLD and right-dislocations; see below the discussion of *recall* and *precision*).

(1) a. *Lo* *vi* *a* *Juan*.[5]
3SG-M_{CL} I.saw A Juan
'I saw (him) Juan.'

b. *Las* *compré* *las* *revistas.*
3PL-F_{CL} I.bought the.PL-F magazines-PL-F
'I bought (them) the magazines.'

CLD, and in particular DOCLD, poses important challenges for syntactic theory. First, all generative theories assume that a verb's valency determines exactly the number of arguments that can appear in its clause. In CLD, this restriction is apparently violated. This has led to much work in syntax proper investigating the difference between agreement and doubling (i.e. if the clitic in CLD is an agreement marker, there is no violation of valency), and the structure of CLD configurations (e.g. big DPs (Torrego 1992, 1995; Uriagereka 1995), clitic voice projections (Sportiche 1996)), as well as work in the syntax/semantics and syntax/pragmatics interfaces (e.g. what semantic/pragmatic factors constrain or explain the appearance of DOCLD).

Yet, even though occurrences of DOCLD are very noticeable (especially for speakers of a Spanish variety with more restricted doubling), the construction is in

5. Glosses will be as transparent as possible, except for: third person DO clitics which will be glossed 3SG-M_{CL}, 3SG-F_{CL}, 3PL-M_{CL}, and 3PL-F_{CL}; "personal *a*", which will be glossed A; and the clitic *se,* glossed SE when it is a lexical clitic, integral part of an "inherently reflexive" verb. All examples are cited exactly as they appeared on the web, including misspellings and other errors.

fact quite rare. This is the case even in Rioplatense, one of the dialects that are most permissive of DOCLD (see, e.g. Suñer 1988). For example, Estigarribia (2006) extracted all transitive clauses from Ligatto (1996), and Fontanarrosa (1995b, 1995a). Of a total of 796 transitive clauses, 579 had no DO clitic, 176 had a DO clitic but no double, and only 40 (5%) had any doubling, either DOCLD or clitic-left dislocation. Only 24 transitive clauses had a postverbal doubled NP (3%). Similarly, Belloro (2007) examined the corpus in Barrenechea (1987), which consists of 33 interactions totaling roughly 24 hours and about 250,000 words and found only 119 DOCLD cases. Interestingly, most studies of DOCLD have no reliable estimates for the relative frequency of different subtypes of DOCLD. For example, it is often claimed that Rioplatense DOCLD is more common with animates than with inanimates. Yet, Estigarribia (2006) showed that this is not necessarily true, since his corpus contains as many cases of doubling of inanimate DOs than of animate ones. It is also generally accepted that DOCLD in Rioplatense requires a specific NP as double. Again, Estigarribia (2006) shows that non-specific NPs can occasionally be doubled, perhaps under a partitive interpretation. Moreover, it is clear that DOCLD interacts in some way with information/cognitive status, be it with topicality, or accessibility (Belloro 2007, 2011, 2012), or with pragmatic presuppositionality (Mazzuchino 2013), but, as Mazzuchino herself notes, the communicative functions of DOCLD are still unclear.

Finally, recent research has uncovered the possibility of "triplications" with clitics, a construction that seemingly combines a clitic left dislocation with CLD (Estigarribia 2013, 2014, forthcoming; López 2009; Suñer 2006). However, this construction is extremely rare. It is not found in any of the corpora cited by Belloro (2007; p. c.) or Estigarribia (2005, 2006). Estigarribia (2014) provides a few naturally occurring examples, but, clearly, examination of a very large corpus of naturalistic speech or speech-like web texts is required to understand under which syntactic, semantic, and pragmatic conditions this construction appears.

These observations highlight the need for great amounts of naturalistic data to test some of the claims made about DOCLD, in particular, claims that the lack of some semantic or pragmatic property (e.g. specificity (Suñer 1988), principal-filterhood (Gutiérrez-Rexach 1999), accessibility (Belloro 2007), presuppositionality (Mazzuchino 2013)) is incompatible with DOCLD. Answers to these questions require not only native speaker judgments but also extensive corpus data on which native speaker intuitions can be brought to bear. To see why this is the case, take Mazzucchino's (2013) proposal that DOCLD is constrained by pragmatic presuppositionality. Clearly, examples where DOCLD occurs in the absence of pragmatic presuppositionality must be rare and/or require contexts of usage that are themselves rare. In that case, native speakers judging DOCLD examples without presuppositionality are expected to have difficulties creating contexts to accommodate the

examples, and, hence, to judge them ungrammatical. This is particularly serious in cases of a change in progress, where a construction slowly expands its occurrence in particular pragmatic or discourse contexts. In general, (inductively) demonstrating that a construction is impossible in conjunction with any factor X requires enormous amounts of evidence (this is known in language acquisition as the "negative evidence" problem). Hence, large numbers of instances in a sample as unbiased as possible (i.e. not disfavoring the identification of particular subsets of DOCLD) are needed. The WWW is a potentially unrivalled source in this light.

Hence, we propose that to investigate natural occurrences of DOCLD in large amounts of data requires both obtaining WWW data and developing NLP tools to identify the construction given an electronic text. Any NLP algorithm has to (1) be able to be efficiently applied to very large amounts of data (an issue we do not directly address in this paper) and (2) avoid as much as possible missing instances of the construction (i.e. false negatives), a requirement we discuss in the following section when we introduce the concepts of *precision* and *recall*.

3. Precision and recall

Precision is defined as the fraction of retrieved instances or "hits" that are actually examples of the target construction (also called *positive predictive value*). *Recall*, on the other hand, is the fraction of instances of the target phenomenon in a sample that are actually retrieved (often called also *sensitivity* or *true positive rate*). In Figure 1, DOCLD cases would be to the left of the line, retrieved items ("hits") in the oval, dark grey areas are errors (false negatives outside the oval, false positives inside).

If a test returns very few false positives (in our case, if it does not often happen that we retrieve non-DOCLD cases), in general it will have high precision. However, this can happen at the expense of an increase in "missed" cases or false negatives. If a test returns a large majority of examples of the target construction, then it will have high recall (even though it might also return a large number of false negatives).

We propose that recall is much more important than precision in the search for infrequent syntactic constructions. This is not a common requirement in web-based corpus linguistics, which tends to favor precision over recall given the vastness of the corpus (Baroni & Kilgarriff 2006; Gatto 2014). To illustrate this point, suppose that a search of 500,000 words retrieves a putative 500 cases of DOCLD. First, it is much easier for a human to weed out false positives from this reduced 500 sentence corpus, than it would be for a human to go back to the original corpus and try to find false negative cases. Second, as we explained above, it is important to miss as

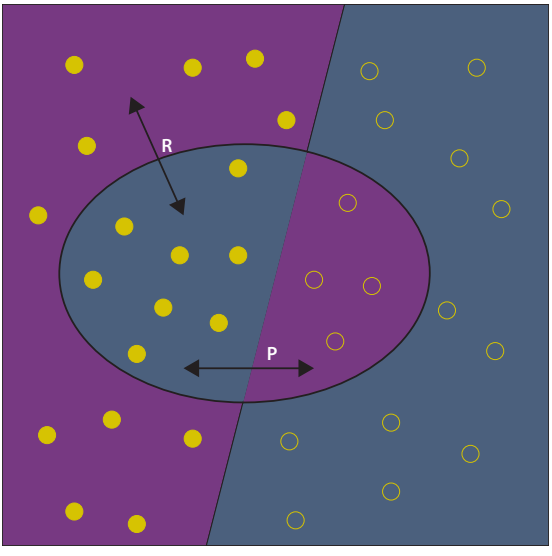


Figure 1. A visual depiction of R(ecall) and P(recision)

few instances of DOCLD as possible, to minimize the chance of missing rare but theoretically illuminating examples.

Recall can be negatively affected by several factors, among which the introduction of errors when preprocessing with extant NLP tools or a corpus' native tools (see below). In the next sections we look at some limitations of extant corpora and NLP tools, when it comes to finding large numbers of instances of DOCLD. Recall affected by low quality of the original data is discusses in Section 6 below.

4. Limitations of off-the-shelf tools (corpora and parsers)

Thanks to the advances made in Natural Language Processing in the last 20 years, we have now several good corpora for Spanish, among which we find Davies' *Corpus del Español* (2002), Subirats and Ortega's *Corpus del Español Actual* (CEA; 2014), or the Real Academia Española's *Corpus de Referencia del Español Actual* (CREA; 2014). These corpora are searchable using regular expressions on lexical items or parts of speech (PoS), which makes them very powerful.

Yet, established corpora also have disadvantages. The first one is their size and composition. The Corpus del Español includes 100 million words, the Corpus del Español Actual 540 million, the CREA 170 million wordforms. Keller and Lapata (2003) already cautioned that for sparse data, constructed corpora are not big enough. Since a large part of texts found in these corpora are from written sources

(where DOCLD is not as frequent), the size of the corpus where DOCLD has a reasonably high chance of appearing is much lower than the total corpus size. Moreover, these corpora conflate dialects where DOCLD is relatively very frequent (e.g. Rioplatense), with dialects in which it is not (e.g. Peninsular).

The second disadvantage is that the exhaustivity of the results we would like for DOCLD is not guaranteed for two reasons. First, the interface itself limits in many cases not only how many hits are displayed but also how they are displayed (e.g. how big a context is provided for a hit). Second, the corpora are already tagged, and this raises the possibility of errors of identification (false negatives) introduced by inaccuracies in corpus tags. For example, a corpus may have 95% global tagging accuracy, but not perform optimally for a given particular construction/part of speech (see as well our discussion of the utility of parsers below).

A further limitation of curated corpora is that their texts are fixed: once you have searched the corpus and extracted all examples, the corpus is exhausted. It is true that corpora can be enlarged, but this usually requires further steps of clean-up and tagging. The WWW is self-updating.

A last limitation, perhaps the main disadvantage of corpora for this project, is that syntactic searches in particular are difficult. None of these extant corpora have syntactic annotations that would facilitate searching for syntactic constructions. While this is also true of web data, this is precisely the limitation this paper tackles.

What we envision as an ideal solution for the problem of identifying syntactic constructions on the WWW is to be able to construct automated tools that are geared specifically to each given goal. An ideal (perhaps too ideal) NLP tool would scour the web for Spanish texts[6] and identify CLD with no human intervention, thereby bypassing the size and updating limitations of corpora. This is a different strategy that complements much work in NLP that attempts to build multi-purpose corpora and tools (e.g. parsers).

5. Pattern identification vs. parsing

Identification of syntactic constructions in texts is in some sense the reverse problem from (syntactic) parsing. In parsing, one is interested in producing a full or partial syntactic representation of a corpus of sentences. Here, we are only interested in identifying instances of a given, well-defined construction (pattern identification). Our proposal is that this requires a carefully thought-out combination of preexisting tools for pre-processing and specific code for the construction at hand.

---

6. An admittedly difficult problem not addressed in this paper.

Targeting a particular syntactic construction for identification (a type of pattern matching problem) as opposed to parsing a corpus is a problem tackled, for example, by Dufter (2009). His goal was to study the counterparts of English *it*-clefts in German, French, Italian, Spanish, and Portuguese. Crucially, Dufter's corpus is EUROPARL, a parallel corpus of European Parliament Proceedings. This allowed him to search for *it*-clefts in English, which are very easily defined using regular expressions on lexical items (not PoS).[7] Even in this relatively simple case, Dufter (2009: 87) cautions that "[t]o be sure, the query string […] will fail to yield the totality of cleft candidates." Regular expressions cannot deal with all possible variations in the structure of clefts.

This is very problematic for DOCLD, since it is not as well-defined lexically as *it*-clefts. DOCLD is only partially defined by PoS, since the Spanish DO clitics are mostly identical to the Spanish articles, with the latter being much more frequent. Any lexical search for *lo/la/los/las* will yield a staggering number of false positives. Furthermore, PoS-matching is not enough. Even though CLD always contains one item of the set of 4 Spanish DO clitics, the set of possible "associate" NPs is infinite and no single PoS is always present. This is because there are dependencies between the PoS: for example, a Det can be missing if a proper noun heads the NP.

(3) a. *Lo vi a tu hermano* (*a*, N; but no Det, no Adj)
3SG-M$_{CL}$ I.saw A your brother
'I saw (him) your brother.'

b. *Lo vi el libro* (Det, N; but no *a*, no Adj)
3SG-M$_{CL}$ I.saw the book
'I saw (it) the book.'

c. *Lo vi el blanco* (Det, Adj; but no *a*, no N)
3SG-M$_{CL}$ I.saw the white
'I saw (it) the white one.'

d. *Los vi a todos* (*a,* Quant; but no Det, no N, no Adj)
3PL-M$_{CL}$ I.saw A all-PL-M

e. *Los vi los que me dijiste* (Det, *que*-headed clause; no N, no Adj)
3PL-M$_{CL}$ I.saw the.PL-M that me you.said
'I saw (them) the ones you told me.'

Hence, although the CEA corpus, for example, allows PoS searches with regular expressions, specifying a successful regex search is cumbersome and inefficient.

7. The structure of *it*-clefts is "*It is/was* + Cleft + relative/Ø + embedded clause". It can be captured by the search string "It|it" "is|was []*that|which|who|whom|when|where|whose|what|how".

Obtaining a parse of sentences with DO clitics can potentially hinder recognition more than help it. A crucial caveat against extant parsers is that their performance is evaluated globally (i.e. measuring parsing accuracy in a given corpus with respect to some gold standard parse of that corpus). Globally accurate parsers, however, can be inaccurate for specific structures. For example, we parsed the constructed sentence (4a) with FreeLing's 3.1 online dependency parser.

(4) a. *Lo* *vi* *[$_{PP}$ en el lugar que me dijiste] a Juan*
3SG-M$_{CL}$ I.saw in the place that me you.told A Juan
'I saw Juan in the place you told me.'

b. **Lo vi [$_{PP}$ en el lugar que me dijiste a Juan]*
'*I saw it/him in the place you told Juan me.'

Example (4a) has an intervening PP ending in a verbform (*dijiste*) which often takes an indirect object headed by *a*. Predictably, the parser mis-attached *a Juan*, actually the DO of the matrix verb *vi*, to *dijiste* as an argument in the embedded clause inside the PP (as shown in 4b). If this parser information is used as a preprocessing step in identifying DOCLD, it will lead to a false negative. It is hard to predict where exactly mis-parses are going to occur.[8] For these reasons, we decided to attempt to build a recognizer without previous parsing.

6. Curated vs. raw text

One last very important factor when attempting to mine web data is how noisy the source is. A particular problem is that the kinds of web texts were DOCLD is most likely to occur (comments in blogs, Facebook, as opposed to more carefully composed text such as blog entries or literary texts) are the noisier kinds of texts, where one would expect to find more misspellings, typos, and lack of punctuation.

We call *raw text* any text presented in the exact form as it appears on the web (minus encoding changes, for example from ASCII or ISO-8859-1 to UTF-8). *Curated text* is text that has been edited for typos, misspellings, and punctuation. We differentiate it from *edited text* that can also have been edited for content, clarity, and involves semantic editing (e.g. ambiguity resolution). Prior data clean-up by

8. Moreover, parsing might not even be a good idea in this case: "[U]nlike programming languages, natural languages are much more about the semantics than about the syntax, so you could be much better off skipping the learning curves of existing parsing tools, going with a home-brewed (top-down, backtracking, unlimited lookahead) lexical analyzer and parser, and spending the bulk of your time writing the code that figures out what a parsed, but ambiguous, natural-language sentence means" (Añez 2011).

a human, while desirable, is very labor intensive and defeats the purpose of automating the process of identification of DOCLD (Gatto 2014). We will see that the performance of our algorithm is very different on curated vs. raw text.

7. Our strategy

We set about designing and programming an explicit, rule-based algorithm for recognition of DOCLD. We did not choose to train a machine learning algorithm because we wanted to understand the way the code classifies DOCLD from non-DOCLD as well as possible in a declarative manner, to better understand recognizer errors. Learning algorithms can show patterns of bias (consistent wrong output for a given input) or variance (variable outputs for a given input) that are hard to foresee (James et al. 2013).

Our algorithm performs "pattern matching", that is, it looks for exact matches of particular syntactic configurations in the data. This is opposed to pattern recognition algorithms that provide a "statistically most likely" classification of any given input. Furthermore, our algorithm is a recognizer, not a classifier, since it only outputs positive cases, and does not classify every input sentence as containing CLD or not.

We strove to minimize preprocessing of the data with existing NLP tools that may introduce noise and lead to higher false negatives (errors introduced by these tools would propagate). Since searching by PoS is essential (see Section 5 above), we chose to preprocess our data with TreeTagger (Schmid 1994, 1995), a very widely used tool (Gatto 2014) with parameter files for 16 different languages. Additionally, the developer kindly agreed to modify his tagging parameters to increase accuracy in tagging DO clitics, if we provided him with new training data. This step turned out to be necessary since at the outset many DO clitics were mistagged as articles, which increased significantly the number of false negatives in our procedure. We, therefore, hand-coded constructed Rioplatense examples containing a high number of DO clitics to improve the performance of the tagger with respect to that particular PoS. We also tagged manually six short stories in Rioplatense, totaling 18,809 words.[9] The developer retrained his model on these new data and we could ascertain that the new TreeTagger version was perfectly accurate on our data for this paper.

TreeTagger outputs three columns in text format. The first column shows the word tokens in the original text, the second column the tags (third-person clitics,

9. All from Rioplatense author Roberto Fontanarrosa (titles: Cenizas, El pichón de Cristo, Uno nunca sabe, La columna tecnológica, Número 5 de Peñarol).

relevantly, are identified with the PPC tag),[10] the third columns the lemmas. This is the input to our code.

Since we also strove to maximize recall, even at the expense of accuracy, our code is very conservative in excluding sentences. It tends to consider as NP doubles all strings that match somewhat lax conditions (see below for a description of the algorithm). It is also greedy, in that as soon as it identifies a candidate NP double, it retrieves it, without continuing looking for other parts of the sentence that are more likely to be matches.

7.1 Description of pattern matching algorithm (CLDFinder)

Given a sentence containing a direct object clitic, CLDFinder tries to find possible NP matches within this sentence, while trying (very conservatively) to exclude intervening PPs or parentheticals. Recognizing possible NP doubles amounts to recognizing NPs in object position. For this paper, we attempted to create a heuristic that would perform this task, by having the code look for the possible different NP structures disjunctively: either you find *a* or you don't; either you find a specifier or you don't, etc.[11]

The way CLDFinder proceeds is as follows. It first finds all sentences containing instances of DO clitics in the previously PoS tagged text, and discards the rest. For each sentence with a DO clitic, it identifies its position and then looks rightward for candidate NPs for doubling. Candidate doubles are identified when CLDFinder finds either <a PREP a> or a possible specifier for NP (any word tagged ART, CARD, DM, etc.). If CLDFinder detects first punctuation that would signal the end of a sentence, it discards the current sentence and goes on to the next. If CLDFinder detects first a comma/dash/opening parenthesis, it considers it "intervening" material. In this case, CLDFinder proceeds to discard all material until the next comma/dash/closing parenthesis. If it finds first a preposition other than "a", CLDFinder skips one specifier and/or one word tagged "noun" to the right and repeats the search for the beginning of a NP after this skipped material. When it finds <a PREP a> or a possible specifier for NP it looks for a noun, before finding any "intervening" material. If it finds a noun, it outputs a match. If it does not, and it finds "intervening" material or the end of the sentence, it discards the sentence and goes on to the next.

---

10. We focused here only on identifying DOCLD with third person DOs (the most interesting case from a theoretical perspective).

11. We discuss the possibility of using instead an extant *chunker* (i.e. doing shallow parsing) in the conclusion.

As we can see, the code is purposefully designed to be very conservative in rejecting sentences or excluding material from matching. For example, if it finds a preposition other than "a", it will exclude exactly ONE specifier and ONE noun at most, even though nouns can be coordinated as complements of a preposition. It will also return as positives cases of right-dislocation or cases when a candidate double is found in a different clause in the same sentence.

8. Results

8.1 Edited, curated text

We applied the CLDFinder procedure to a short story ("Uno nunca sabe", 4427 words) by Argentinian (Rioplatense) writer Roberto Fontanarrosa (accessed at http://www.taringa.net/posts/humor/6264604/Cuentos-de-Fontanarrosa---Uno-nunca-sabe.html, *1 October 2014*). Fontanarrosa's short stories are written in the vernacular of the Argentine city of Rosario, Santa Fe, itself part of the Rioplatense area. Many of them are based on the dialogue between the characters, minimizing narration (as is the case for "Uno nunca sabe", as well as for Fontanarrosa 1995b; and Fontanarrosa 1995a). Table 1 shows our performance measures on curated text. We provide F1 and F2 scores for completeness. They both measure a classification test's accuracy from 0 to 1 by calculating the trade-off between precision and recall. F1 is a balanced score where precision and recall have equal weight; F2 gives recall twice as much weight as precision (Alonso et al. 2008).

Table 1. Performance measures of CLDFinder on curated text

	Tokens	**Recall**	**Precision**	**F1**	**F2**
Short Story (curated text)	8	100%	67%	.8	.91

This short story had 8 tokens of CLD. Recall was 100%, meaning that we captured all CLD cases. Precision, however, was 67%, meaning that one third of hits (4 cases) CLDFinder retrieved were actually not CLD cases. The performance is highly satisfactory since we have no false negatives, and false positives are kept to a minimum.

Errors leading to false positives belonged to three classes: tagging errors, semantics-dependent errors, and code errors. *Tagging errors* (1 case) are errors introduced by the tagger during preprocessing (5; false matches indicated with dashed underlining).

(5) *A esa mina te la ibas a pirobar*
A that.SG-F girl SE 3SG-F$_{CL}$ you.went to fuck
'You were going to fuck that girl.'

Pirobar is a Rioplatense lunfardo ('slang') word meaning 'to fuck'. However, it was not recognized by the tagger as such. It was instead tagged as a common noun. This allowed CLDFinder to match it to the clitic, in a configuration CL + DO(personal "a").

Two more false positives were the result of a *semantics-dependent error*. These are cases where the correct syntactic analysis can only be assigned by retrieving the participant role of a constituent that makes sense in the context. In (6), *a la mesa* is intended as a locative, but on the surface, it could have been a NP double. In Rioplatense Spanish, personal "a" is possible with some inanimates in contexts of high transitivity. In this bringing event, the patient is highly affected and individuated, being concrete, singular, countable, and definite (Hopper & Thompson 1980). This means that *a la mesa* can have a patient interpretation and be mapped as DO of the verb, as in (7).

(6) *Ningún amigo te la trajo a la mesa con vos*
no friend to.you 3SG-F$_{CL}$ brought to the table with you
'No friend brought her to your table.'

(7) *¿Te la trajo a la mesa el carpintero?*
to.you 3SG-F$_{CL}$ he.brought A the table the carpenter
'Did the carpenter bring you the table?'

The remaining error is specific to the way the code looks for CLD matches; we call these *code errors*.

(8) *[se la pasa mirándome] y [jamás se ha atrevido a decirme nada].*
SE 3SG-F$_{CL}$ passes looking = me and never SE has dared to tell = me nothing
'He is always staring at me and he's never dared tell me anything.'

Example (8) combines a tagging error with a code-specific error in the identification of clause limits. First, *la* here is a lexical clitic, part of the meaning of Rioplatense *pasársela* 'to constantly, repeatedly do something'. Since this clitic is not referential, it cannot be part of a CLD configuration. But the tagger cannot differentiate between referential and non-referential clitics. However, in this case, even if the

clitic had been referential, there was no coreferring DO in the same clause (clause boundaries are marked with added square brackets for clarity).[12]

8.2 Raw text from the web

In order to gauge CLDFinder's performance with raw web texts, we extracted a page from the (now defunct) Argentinian soccer blog "Vale Chumbar/Muy Fútbol". This multi-author, interactive blog, consisted of posts about Argentinian and world soccer, and allowed users to post comments. The discussion being of a sports nature, comments were written in a very vernacular, colloquial, almost dialogical style, and could be rather passionate (some legendary feuds among regular commenters arose, and subsisted over the many years the blog was functioning).

We chose one of the pages with most comments overall (342 comments, 2814 words). The post's title was "Ponele la frase a Goleo" (Goleo, a lion, was one of the official mascots of the 2006 FIFA Soccer World Cup in Germany, the other mascot being his sidekick Pille, a football). The author of the post asked commenters to contribute phrases for the pictures Goleo appeared in. Ribald jokes were, of course, rife. The length of comments varied from 1 word to 25 words.

CLDFinder did substantially more poorly on this text. The text contains six instances of CLD. Table 2 shows that of the 8 items retrieved, only three were CLD cases (precision 37.5%), while 3 actual instances of CLD were missed (recall 50%).

Table 2. Performance measures of CLDFinder on raw web text

	Tokens	Recall	Precision	F1	F2
Webpage (raw text)	6	50%	37.5%	.43	.39

8.2.1 *False positives*

Errors yielding false positives were again of the same three classes: tagging errors (1 occurrence), semantics-dependent errors (2 occurrences), and code errors (2 occurrences).

In (9), the phrasal preposition *junto a* 'close to, together with' was mis-tagged as a sequence composed of the adjective *junto*, followed by the preposition *a*. CLDFinder treated this as a possible match.

12. Even though non-referential DO constituents such as *nada* 'nothing' presumably can never be doubled, in a very conservative approach like ours, one certainly wants to allow matching them.

(9) *No me extrañaría verlo laburando en lo de tinelli*
no me would.surprise see = him working in the of Tinelli
junto al gordo Larry
together.with A.the.SG-M fat Larry
'It wouldn't surprise me to see him working on Tinelli's TV show together with Fat Larry.'

Example (10) shows a semantics-dependent error. The sentence is not recognized as an instance of the idiom *dejar algo a criterio de alguien* ('leaving something to someone's discretion'), where *a criterio de alguien* is a prepositional complement fixed in the idiom. The same surface structure can be a CLD case with a different common noun (11). In (12), CLDFinder matched *a 4to bat* (a commenter's username) to *la*, even though from context it is clear that *la* refers to a referent in the previous context, and the *a 4to bat* is coreferential with the dative clitic *se*. However, the surface string (out of context) is compatible with the CLD interpretation (for example, if *4to bat* referred to a feminine discourse referent).

(10) *Lo dejo a tu criterio*
3SG-M_{CL} I.leave to your criterion
'I leave it up to you.'

(11) *Lo dejo a tu libro*
3SG-M_{CL} I.leave A your book
'I leave (it) your book.'

(12) *Se la deje a 4to bat*
to.him 3SG-F_{CL} I.left to "4th bat"
'I left it to/for "cuarto bat".'

Finally, there were two errors involving failures of the CLDFinder code.

(13) *[Lo admito…] [soy el gauchito del 78]*
3SG-M_{CL} I.admit… I.am the little.gaucho of.the '78
'I admit it… I am the little gaucho of '78.'

(14) *Porque se lo cogio [$_{NP}$ todo el pabellón].*
because SE 3SG-M_{CL} fucked all the pavilion
'Because the whole pavilion fucked him.'

Example (13) shows failure of CLD Finder to recognize a clausal boundary. In (14), CLDFinder matched the clitic with *todo*. While *todo* allows such matches often (in many varieties of Spanish DOCLD with *todo* is quite common), in this case *todo* is a quantifier in the subject NP. Yet, this error is also partly a semantics-based error.

Note that with two lexical replacements, the same surface structure can have two CLD analyses.

(15) *Porque se lo leyó [$_{NP}$ todo el libro].*
because SE 3SG-M$_{CL}$ read all the book
'Because s/he read it, the whole book.'

(16) *Porque se lo leyó [$_{QP}$ todo] [$_{NP}$ el libro].*
because SE 3SG-M$_{CL}$ read all the book
'Because s/he read it all, the book.'

In (14), we know that *todo el pabellón* is subject from the context of the discussion. So while it is true that CLDFinder was greedy and matched *lo* to the first possible match without looking for alternative matches, the correct structural assignment can only be arrived at by semantic interpretation.

8.2.2 *False negatives*

The three false negatives are a more serious problem. There is one tagging error and two errors caused by typos in the comments. Crucially, there are no code-specific errors.

The tagging error in (17) stems from the tagger's failure to recognize the enclitic = *lo*. This is because the Rioplatense verb form *ponelo* (as opposed to the Standard Spanish *ponlo*) is not correctly recognized as an enclitic-bearing form.

(17) *Ponelo al enzo*
put=3SG-M$_{CL}$ A.the.SG-M Enzo
'Make Enzo play (in the soccer game).'

The other two false negatives are due to misspellings. In (18), *Heinze* is recognized by the tagger as a proper noun, but the commenter did not include "personal a", so CLDFinder did not match it with the clitic. In (19), the commenter omitted the space after "personal a" and the determiner was not recognized. Since determinerless phrases are only allowed to match a clitic if they are proper nouns, and *argentinos* is not a proper noun (number mismatch is ignored), this resulted in no match.

(18) *Lo vieron Heinze en sin cassete??*
3SG-M$_{CL}$ you.saw Heinze on without tape
'Did you see Heinze on "Without tape" (TV show)?'

(19) *Como lo cagamos alos argentinos*
how 3SG-M$_{CL}$ we.shat.on A + the argentines
'We really screwed the Argentines.'

9. Discussion

Overall, CLDFinder performs very well on curated text (texts free of spelling and punctuation mistakes), but its performance decreases significantly on raw text. Therefore, while these results are very encouraging, significant work remains to be done.

The performance of CLDFinder depends crucially on factors external and internal to the code. CLDFinder is very conservative (it tends to accept as many possible matches as it can) and greedy (once it finds a candidate NP double, it matches it without looking more closely or for more likely matches). One factor internal to the code is the lack of appropriate clause boundary recognition, which creates false positives by allowing matches of a clitic to a NP in a subsequent clause. However, while this a serious shortcoming, false positives can be more easily discarded by a human coder post hoc, than false negatives identified.[13]

The external factors are those that involve the characteristics of the text and any pre-processing steps needed to prepare the input for CLDFinder. Inconsistent punctuation, typos, and misspellings are a major problem in the automatic analysis of raw web data. One possible solution is to include another preprocessing step with automatically cleans the data, before tagging. Moreover, errors introduced by the PoS tagger even in curated text can lead to false negatives. These errors are rather serious, since we would like to avoid false negatives at all costs. Therefore, immediate steps for improvement should tackle mainly external factors.

The easiest step forward is to provide the tagger with more Rioplatense training data, in particular, to allow TreeTagger to better recognize Rioplatense verb forms with enclitics (Example 17). Another option is to use a Rioplatense dictionary and re-encode specific word forms as forms that are recognized by the tagger (while maintaining the original information for later retrieval).

The tagger's performance can also be improved by, for example, adding a preprocessing step of conservatively retokenizing the text (useful for cases like (19)). However, the benefits of each added preprocessing step with off-the-shelf tools that may not be optimized for the particular problem at hand need to be carefully weighed against the likelihood of their introducing more error.

To conclude, we have demonstrated the usefulness of developing "home-brewed" specific NLP tools to analyze web texts in syntax research. We hope that this approach can inspire others to continue seeking ways to complement the use

13. We agree with a reviewer that the larger the database, the more time consuming it will become for a human coder to weed out false positives. Nevertheless, it is at least in principle more feasible than to go back to the original data set and try to find false negatives, which completely defeats the purpose of our code in the first place. Clearly, a larger database is needed to test these issues.

of established, curated corpora at the intersection of corpus and computational linguistics.

References

Alonso, Jaime, Juan José del Coz, Jorge Díez, Oscar Luaces, and Antonio Bahamonde. 2008. "Learning to Predict One or More Ranks in Ordinal Regression Tasks." In *Machine Learning and Knowledge Discovery in Databases: European Conference, Antwerp, Belgium, September 15–19, 2008, Proceedings*, edited by Walter Daelemans, Bart Goethals, and Katharina Morik, 39–54. Berlin Heidelberg: Springer Science & Business Media. doi: 10.1007/978-3-540-87479-9_21

Añez, Juancarlo. 2011. "Reply to 'Efficient Context-Free Grammar Parser, Preferably Python-Friendly.'" *Stackoverflow*. http://stackoverflow.com/questions/4543008/efficient-context-free-grammar-parser-preferably-python-friendly.

Baker, Paul. 2010. *Sociolinguistics and Corpus Linguistics*. Edinburgh University Press.

Baroni, Marco, and Adam Kilgarriff. 2006. "Large Linguistically-Processed Web Corpora for Multiple Languages." *EACL 2006–11th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference*, 87.

Barrenechea, Ana María. 1987. *El habla culta de la ciudad de Buenos Aires: materiales para su estudio*. edited by Instituto de Filología y Literaturas Hispánicas "Dr. Amado Alonso." 2 vols. Buenos Aires: Universidad nacional de Buenos Aires, Facultad de filosofía y letras.

Belloro, Valeria A. 2007. "Spanish Clitic Doubling: A Study of the Syntax-Pragmatics Interface." PhD dissertation, Buffalo, NY: State University of New York at Buffalo. http://search.proquest.com/llba/docview/85694379/137F58F7E4577BC9AC2/1?accountid=14244.

Belloro, Valeria A. 2011. "Dislocaciones Y Doblados: Entre La Concordancia Anafórica Y La Gramatical." *Hechos Y Proyecciones Del Lenguaje* 20: 127–49.

Belloro, Valeria A. 2012. "Encoding Information Structure via Object Agreement in Spanish Interactions." In *Proceedings of BLS* 34, 391–402. Berkeley, CA.

Davies, Mark. 2002. "Corpus Del Español." Corpus of Spanish. http://www.corpusdelespanol.org/.

Dufter, Andreas. 2009. "Clefting and Discourse Organization: Comparing Germanic and Romance." In *Focus and Background in Romance Languages*, edited by Andreas Dufter and Daniel Jacob, 83–121. Amsterdam: John Benjamins Publishing. https://www.academia.edu/928802/Clefting_and_discourse_organization_Comparing_Germanic_and_Romance. doi: 10.1075/slcs.112.05duf

Estigarribia, Bruno. 2005. "Direct Object Clitic Doubling in OT-LFG: A New Look at Rioplatense Spanish." In *The Proceedings of the LFG '05 Conference*, edited by Miriam Butt and Tracy Holloway King. University of Bergen, Norway. http://csli-publications.stanford.edu/LFG/10/lfg05estigarribia.pdf.

Estigarribia, Bruno. 2006. "Why Clitic Doubling? A Functional Analysis for Rioplatense Spanish." In *Selected Proceedings of the 8th Hispanic Linguistics Symposium*, edited by Timothy L. Face and Carol A. Klee, 123–36. Somerville, MA: Cascadilla Proceedings Project.

Estigarribia, Bruno. 2013. "Rioplatense Spanish Clitic Doubling and 'Tripling' in Lexical-Functional Grammar." In *Selected Proceedings of the 15th Hispanic Linguistics Symposium*, edited by Chad Howe, Sarah E. Blackwell, and Margaret Lubbers Quesada, 297–309. University of Georgia, Athens: Cascadilla Proceedings Project.

Estigarribia, Bruno. 2014. "La estructura informacional en la triplicación con clíticos del español rioplatense." *Signo y Seña | Revista del Instituto de Lingüística*, no. 25: 105–132.

Estigarribia, Bruno. Forthcoming. "The semantics of Spanish Clitic Left-Dislocations with epithets." To appear in *Probus*.

Fletcher, William H. 2012. "Corpus Analysis of the World Wide Web." In *The Encyclopedia of Applied Linguistics*, n.p. Blackwell Publishing Ltd. http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal0254/abstract. doi:10.1002/9781405198431.wbeal0254

Fontanarrosa, Roberto. 1995a. "Beto." In *La Mesa De Los Galanes y otros cuentos*, 42–50. Buenos Aires: Ediciones De La Flor.

Fontanarrosa, Roberto. 1995b. "Periodismo investigativo." In *La Mesa De Los Galanes y otros cuentos*, 7–26. Buenos Aires: Ediciones De La Flor.

Gatto, Maristella. 2014. *Web As Corpus: Theory and Practice*. A&C Black.

Gimpel, K., N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and N. A. Smith. 2011. "Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments." In *Proc. of ACL*.

Gries, Stefan Thomas. 2009. *Quantitative Corpus Linguistics with R: A Practical Introduction*. 1st ed. Routledge. doi:10.1515/9783110216042

Gries, Stefan Thomas, Stefanie Wulff, and Mark Davies, eds. 2009. *Corpus-Linguistic Applications: Current Studies, New Directions*. Rodopi.

Gutiérrez-Rexach, Javier. 1999. "The Formal Semantics of Clitic Doubling." *Journal of Semantics* 16 (4): 315–380. doi:10.1093/jos/16.4.315

Hopper, Paul J., and Sandra A. Thompson. 1980. "Transitivity in Grammar and Discourse." *Language* 56 (2): 251–299. doi:10.2307/413757

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning (with Applications in R)*. Springer Texts in Statistics 417. New York-Heidelberg-Dordrecht-London: Springer. http://www.springer.com/statistics/statistical+theory+and+methods/book/978-1-4614-7137-0. doi:10.1007/978-1-4614-7138-7

Keller, Frank, and Mirella Lapata. 2003. "Using the Web to Obtain Frequencies for Unseen Bigrams." *Computational Linguistics* 29 (3): 459–484. doi:10.1162/089120103322711604.

Kilgarriff, Adam, and Gregory Grefenstette. 2003. "Introduction to the Special Issue on the Web As Corpus." *Computational Linguistics* 29 (3): 333–347. doi:10.1162/089120103322711569.

Ligatto, Dolorès. 1996. *Matériau pour l'étude de l'espagnol parlé: la variante argentine*. Presses Univ. Limoges.

López, Luis. 2009. *A Derivational Syntax for Information Structure*. Oxford-New York: Oxford University Press. doi:10.1093/acprof:oso/9780199557400.001.0001

Mazzuchino, María Gabriela. 2013. "El doblado de acusativo en el español de Argentina: definitud, especificidad, presuposicionalidad y otras nociones conexas." *Lengua y Habla* 17 (0): 118–152.

Real Academia Española. 2014. "Corpus de Referencia Del Español Actual. Banco de Datos (CREA) [en Línea]." Corpus of Spanish. http://www.rae.es.

Russell, Matthew A. 2013. *Mining the Social Web*. Second edition. Sebastopol, CA: O'Reilly Media. http://shop.oreilly.com/product/0636920030195.do.

Schmid, Helmut. 1994. *Probabilistic Part-of-Speech Tagging Using Decision Trees*.

Schmid, Helmut. 1995. "Improvements in Part-of-Speech Tagging with an Application to German." In *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland.

Sportiche, Dominique. 1996. "Clitic Constructions." In *Phrase Structure and the Lexicon*, edited by Johan Rooryck and Laurie Ann Zaring, 213–276. Dordrecht, Netherlands: Kluwer Academic Publishers. doi:10.1007/978-94-015-8617-7_9

Subirats, Carlos, and Marc Ortega. 2014. "Corpus Del Español Actual (CEA)." Corpus of Spanish. http://sfncorpora.uab.es/CQPweb/cea/.

Suñer, Margarita. 1988. "The Role of Agreement in Clitic-Doubled Constructions." *Natural Language & Linguistic Theory* 6 (3): 391–434. doi:10.1007/BF00133904.

Suñer, Margarita. 2006. "Left Dislocations with and without Epithets." *Probus* 18 (1): 127–158. doi:10.1515/PROBUS.2006.005.

Torrego, Esther. 1992. "Case and Argument Structure." Unpublished manuscript. Boston, University of Massachussets.

Torrego, Esther. 1995. *On the Nature of Clitic Doubling*. http://search.proquest.com.libproxy.lib.unc.edu/llba/docview/85632793/13E384E8D4B6E84ED8D/5?accountid=14244.

Uriagereka, Juan. 1995. "Aspects of the Syntax of Clitic Placement in Western Romance." *Linguistic Inquiry* 26 (1): 79–123.

PART III

Bilingualism and language acquisition

CHAPTER 9

Voice quality transfer in the production of Spanish heritage speakers and English L2 learners of Spanish

Ji Young Kim
University of California, Los Angeles

The present study examines the use of creaky voice in the Spanish of two groups of Spanish-English bilinguals, namely, U.S. heritage speakers of Mexican Spanish and English L2 learners of Spanish. In American English, creaky voice is used utterance-finally, especially among young women, while it has rarely been found in Mexican Spanish. Participants' H1-H2 values in utterance-final position were calculated to measure the degree of creaky voice and the values were compared to those of monolingual speakers of Mexican Spanish. Results showed that while monolinguals did not use creaky voice, both heritage speakers and L2 learners did, especially the female speakers. The findings imply that, like other linguistic features, voice quality can be transferred from one language to another.

Keywords: voice quality transfer, creaky voice, heritage speakers, gender difference

1. Introduction

The present study is motivated by anecdotal reports by Spanish instructors and second language (L2) researchers that creaky voice is used very frequently among English L2 learners of Spanish when they speak Spanish, a phenomenon which is perceived very saliently and sometimes negatively (Kim 2013). However, no empirical research has been done on this phenomenon to answer why and to what extent these speakers use creaky voice in their production of Spanish. Given that creaky voice is often used stylistically in American English (Podesva 2013; Todaka 1994; Wolk et al. 2012; Yuasa 2010), the present study examined creaky voice as a possible feature of transfer from American English to Spanish. Based on research on cross-linguistic phonetic transfer in bilingual speakers (Flege et al. 1995; Flege et al. 1999; Grosjean 1989; Kuhl et al. 2003; Major 1992, among others), the present

DOI 10.1075/rllt.11.09kim

study intends to see whether voice quality (in this case, creaky voice) can also be transferred from the more dominant language to the less dominant one by analyzing the use of creaky voice among two groups of Spanish-English bilingual speakers with different linguistic and cultural background: U.S. heritage speakers of Mexican Spanish and English L2 learners of Spanish. Spanish heritage speakers and English L2 learners of Spanish are similar to each other in that they acquired American English and Spanish, more likely the Mexican variety, and that they are more dominant in English than in Spanish. However, they also differ from each other in that heritage speakers generally acquire Spanish as the first language (L1) and English as the L2, while it is the opposite for L2 leaners. Moreover, given that heritage speakers have stronger ties to their Mexican heritage, they tend to be more exposed to Spanish than L2 learners.

2. Background

2.1 Creaky voice in American English and Mexican Spanish

Creaky voice, also known as "laryngealization" or "vocal fry", is a phonation type that occurs when the vocal folds are tightly approximated, allowing very little room for air escape. It is distinguished from other phonation types such as modal voice (i.e., regular voice) and breathy voice which occurs when the vocal folds are held further apart. Because creaky voice is produced with a constricted glottis, it leads to a longer closed phase of the glottal period and a shorter open phase (Gick et al. 2013; Gordon & Ladefoged 2001; Johnson 2003; Ladefoged 1971; Laver 1980). This is visually differentiated from modal voice in the spectrograms and waveforms in that glottal pulses are further apart and irregularly spaced during creaky voice (see Figure 1). Moreover, creaky voice is produced at the bottom of a speaker's pitch range; thus, it is usually produced with a very low fundamental frequency (Johnson 2003). Creaky voice is associated with various acoustic measures (e.g., lower fundamental frequency, higher jitter, higher shimmer, positive spectral tilt) (Gick et al. 2013; Gordon & Ladefoged 2001; Johnson 2003; Kirk et al. 1993; Wolk et al. 2012), but spectral tilt, especially the relative amplitude of the first two harmonics (H1-H2), is considered to be one of the most reliable cues for both the perception and the production of creaky voice (Keating & Esposito 2006; Kreiman et al. 2010; Johnson 2003).

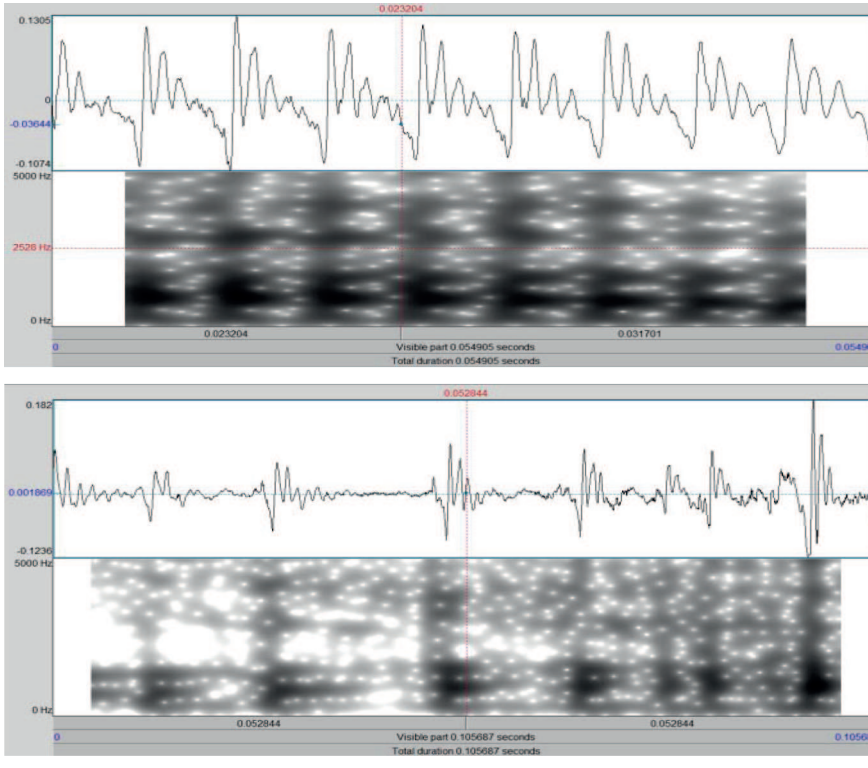


Figure 1. Spectrograms and waveforms of /a/ produced with modal voice (above) and creaky voice (below)

Some languages use creaky voice phonemically to contrast with modal voice and/or breathy voice (e.g., Kwakw'ala: Boas 1947; Jalapa Mazatec: Kirk et al. 1993; Ladefoged & Maddieson 1996; Hupa: Golla 1970). In other languages, such as English and Spanish, in which creaky voice does not have a contrastive function, it is used stylistically (expressing toughness: Mendoza-Denton 2011, Podesva 2013; stance: Dilley et al. 1996, Grivicic & Nilep 2004; commiseration or complaint: Brown & Levinson 1987; fatigue: Colton et al. 2011, Gick et al. 2013).

Given that creaky voice is produced with a very low fundamental frequency, it has been observed more frequently among men due to their larger vocal folds and vocal tract compared to women (Klatt & Klatt 1990; Ohala 1994). Due to the link between male voice and the low pitch that is usually accompanied with creaky voice, in the English-speaking world, creaky voice has traditionally been associated with masculinity and other social meanings connected to it, such as authority and high social status (Esling 1978; Henton & Bladon 1988; Pittam 1987; Stuart-Smith 1999). Typically, female voice has been associated with breathy phonation (Klatt

& Klatt 1990; Pittam 1987; Todaka 1994). However, in the context of American English, creaky voice is used frequently among young female speakers, predominantly at the end of an intonational phrase (IP), which is possibly due to steady drop in fundamental frequency throughout the course of an utterance (Henton & Bladon 1988; Ladefoged 1993; Wolk et al. 2012). The use of creaky voice among female speakers may have originated as a strategy that they adopted "to project a male-like authoritative image" (Coates 1986; Dilley et al. 1996; Podesva 2013; Tajfel 1974; Yuasa 2010), but recently the conventional association between creaky voice and masculinity seems to have weakened (Podesva 2013) because creaky voice is now considered as a socio-linguistic phenomenon, distinctively used by young American women who are educated and upwardly-mobile (Podesva 2013; Todaka 1994; Wolk et al. 2012; Yuasa 2010).[1]

With regard to Spanish, research on the voice quality of native Spanish speakers claims that female speakers in general have a slightly breathier voice than male speakers (Mendoza et al. 1996; Trittin & de Santos y Lleó 1995). Although creaky voice has also been observed in the speech of female Spanish speakers (Armstrong et al. 2015; Morrison & Escudero 2007), possibly as a socio-indexical marker of lack of education and lack of femininity (Armstrong et al. 2015), its use seems to be restricted to Peninsular Spanish speakers, since it has not been observed in other varieties of Spanish (Andrade 2003; Esposito 2006, 2010; Morrison & Escudero 2007). In the case of Mexican Spanish, the variety that is of interest in the present study, Esposito (2006, 2010) found that Mexican Spanish speakers, both male and female, used only modal phonation across different sentence types (e.g. declarative, interrogative, imperative).

2.2 Voice quality transfer among bilingual speakers

According to Gordon and Ladefoged (2001), voice quality is a feature that can be controlled by the speaker according to the language he speaks. If a bilingual speaker speaks two languages of different voice qualities, would she/he be able to keep them apart or will transfer of voice quality occur? Indeed, several studies have shown that voice quality varies across languages and bilingual speakers are able to distinguish them (for French/Dutch: Harmegnies & Landercy 1985; for Catalan/Spanish: Bruyninckx et al. 1994; Harmegnies et al. 1989; for Japanese/English:

1. Since creaky voice is also a characteristic of some female pop stars, it is perceived negatively as a fashion trend sounding "untrustworthy" (Anderson et al. 2014; Quenqua 2012). However, in a matched-guise test, Yuasa (2010) found that, contrary to popular belief, listeners perceived creaky voice as an indexical marker of an educated, urban-oriented, and upwardly-mobile female persona.

Todaka 1994). However, these studies are mostly based on balanced bilinguals and there are few studies that directly address the question of whether transfer of voice quality occurs when one language is more dominant than the other. Esling (1982) and Esling and Wong (1983) reported that it has been noted by several ESL (English as a Second Language) instructors that native Persian speakers tend to speak with breathy voice when they speak English. Given that Persian is characterized as alternating its phonation between breathy and whispery voice in certain linguistic contexts (Heselwood & Mahmoodzade 2007), these learners may have transferred their voice quality from Persian (L1/more dominant language) to English (L2/less dominant language), giving the impression that they speak with breathy phonation. Moreover, in a case study of a female French-English bilingual speaker living in the US, Nichols (2012) found that this speaker produced creaky phonation not only in English, but also in French, more in sentence-final position than in sentence-medial position. Given that creaky voice is not a phenomenon generally found in French (Nichols 2012), the bilingual's use of creaky voice in French implies that this feature may have been transferred to French (L1) from English (L2) which is likely to be the more dominant language. According to Nichols (2012), not only did the bilingual learn English at a young age (age 10), but she used it more frequently than French. Thus, based on the observations above, regardless of the order of acquisition, transfer of voice quality seems to occur from the more dominant language to the less dominant one.

By comparing two groups of Spanish-English bilingual speakers whose dominant language is English, namely, U.S. heritage speakers of Mexican Spanish and English L2 learners of Spanish, the present study examined whether voice quality, more specifically, creaky voice, is transferred from the more dominant language (American English) to the less dominant one (Mexican Spanish). Although creaky voice has been observed in the English of Spanish heritage speakers (Fought 2003; Mendoza-Denton 2011), research has rarely been done on the use of creaky voice in their Spanish and the Spanish spoken by L2 learners. Thus, the present study investigated whether creaky voice is also found in the Spanish of heritage speakers and L2 learners.

3. Methods

3.1 Participants

The present study is part of a larger project. The data were taken from three groups of participants: 9 monolingual native speakers of Mexican Spanish (NS) (5F, 4M), 10 heritage speakers of Mexican Spanish (HS) (7F, 3M), and 17 English L2 learners

of Spanish (L2) (11F, 6M). All the participants matched in age in that they were all college students (age range: 18–26 years). The HSs and the L2s were recruited and tested at a university in the Midwest of the U.S. and they all spoke American English. Both groups of participants were born and raised in the U.S. and reported that they predominantly use English (more than 80% of the time). The HSs were born and raised in homes in which both parents immigrated to the U.S. as adults from different areas in Mexico, mostly from the central-west region, and grew up speaking both Spanish and English. The HSs were considered to be English-dominant, based on three factors: (1) age of acquisition, (2) language use, and (3) language proficiency. All the HSs were early bilinguals in that they acquired both Spanish and English before the age of 5, which is before the period when foreign accent starts appearing if a language is not learned by then (Flege 1992). With respect to language use, the HSs reported that they currently use English (average 81.5%) far more frequently than Spanish (average 18.5%). Regarding overall language proficiency, in a five-point Likert scale, from 1 (poor) to 5 (native-speaker command), the HSs self-rated their Spanish proficiency (average 3.68) lower than their English proficiency (average 4.68). The L2s, on the other hand, were born and raised in a family where only English was spoken and did not learn Spanish until after the age of 9. Similar to the HSs, The L2s reported that they currently use English (average 90.82%) much more frequently than Spanish (average 9.18%) and self-rated their Spanish proficiency (average 2.23) lower than their English proficiency (average 5). Only the L2 data of white male and female speakers were included in the present study because previous research on race and phonation has found that African-American male and female speakers tend to use falsetto phonation, which is produced with tightly adducted and stretched vocal folds, resulting in a high fundamental frequency (Esling 1984), more than white male and female speakers (Podesva 2007, 2013). Lastly, the NSs participated as a control group. All the NSs were monolingual native speakers of Mexican Spanish and they were all recruited and tested at a university located in north-central Mexico. They reported that they use Spanish more than 90% of the time and they did not learn a second language until later in adolescence. None of the participants reported having any speech disorders.

3.2 Materials and procedures

The data consisted of 64 declarative sentences in which the subjects appeared utterance-finally (e.g. *Por la plaza, paso yo.* 'Through the square, I pass'). The materials were designed in such a way to be used as stimuli for a larger project to see whether listeners are able to identify the subject of each sentence, when the subject, located

in utterance-final position, was deleted. In the present study, the voice quality of these utterance-final subjects was analyzed.

The materials were presented in PowerPoint slides on a computer screen and the participants read them out loud. In both the U.S. and Mexico, the productions were audio-recorded in a sound-attenuated booth. In the U.S., the recordings were collected using an AKG C520 head-mounted microphone, which was positioned approximately 2 inches away from the participants' lips, and a Marantz PMD570 solid state recorder. In Mexico, the recordings were collected using an Olympus ME-31 compact gun microphone and an Olympus LS-11 linear PCM recorder. The participants were instructed to maintain a distance of approximately 2 inches from the microphone. All the recordings were done with a sampling rate of 44.1 kHz and a sample size of 16 bits.

3.3 Analysis

In order to control for the effect of intonational contour, tokens with rising intonation at the end of the utterance (87 out of 2304 tokens) were excluded from the analysis. The data were not analyzed categorically (i.e. creaky vs. no creaky), given that a previous study (Kim 2013) had found that categorical data do not adequately reflect different degrees of creaky voice. Thus, in the present study, the use of creaky voice was analyzed in a continuous manner by measuring the relative amplitude (dB) of the first two harmonics of the voice source (H1-H2), which is considered to be one of the most reliable acoustic cues that differentiate phonation types (Keating & Esposito 2006; Kreiman et al. 2010; Johnson 2003). Breathy phonation is produced with strong first harmonic, leading to positive H1-H2 values, while creaky phonation is produced with higher amplitude in the second harmonic than the first harmonic, leading to negative H1-H2 values. The H1-H2 values of modal phonation usually fall between those of breathy and creaky phonations (i.e. around 0 dB). IP-final H1-H2 values were measured as the average of the H1-H2 values of the samples within the last syllable of each utterance. The average of the H1-H2 of the entire IP was also measured to see participants' overall voice quality. This measure was taken to verify whether creaky voice was produced only in IP-final position or whether it is an idiosyncratic speech style of the participants.

The effects of group (NS/HS/L2), gender (F/M), position (IP-final/Entire IP), and the interactions among the fixed factors on H1-H2 values (dB) were analyzed using linear mixed effects modeling with subject and item as random factors. The *lmer()* function in the *lme4* package in R (Baayen 2008) was used for the statistical analysis. The best fitting model according to backwards selection included random intercepts for subject and item with by-item random slope for gender. All the

fixed factors (i.e. group, gender, and position) were centered using contrast-coding. Following Baayen (2008), absolute t-values higher than 2 were considered as statistically significant. For post-hoc pairwise comparisons, the *lsmeans()* function in the *lsmeans* package in R (Baayen 2008) was used.

4. Results

Results showed that there was a significant main effect of position ($\beta = -3.953$, $SE = 0.194$, $t = -20.367$), which suggests that the H1-H2 values were in general lower (i.e., with creakier phonation) in IP-final position when compared to the entire IP (baseline condition). Moreover, significant main effect of group ($\beta = -2.422$, $SE = 1.135$, $t = -2.133$) and significant interaction between position and group ($\beta = -3.426$, $SE = 0.465$, $t = -7.364$) were found for the HSs. This indicates that overall HSs' H1-H2 values were lower than those of the NSs (baseline condition) and the difference between the H1-H2 values in the two positions (i.e., IP-final position and the entire IP) was larger for the HSs than the NSs. Although the L2s showed a similar tendency, this did not reach significance level. The mean H1-H2 values in the two positions showed that HSs' mean IP-final H1-H2 value was slightly lower (−2.77 dB) than that of the L2s (−2.39 dB), while the two groups had similar mean H1-H2 values for the entire IP (HS: 2.43 dB, L2: 2.45 dB), which may have resulted in HSs' significantly larger difference between the H1-H2 values in the two positions when compared to the NSs (IP-final: 1.58 dB, entire IP: 2.21 dB). However, pairwise comparisons of position and group confirmed that the H1-H2 values significantly differed between the two positions not only for the HSs, but also for the L2s ($p < 0.001$), while they were similar to each other for the NSs. When comparing the H1-H2 values between IP-final position and the entire IP in Figure 2, one can see that while NSs' H1-H2 values do not differ significantly between the two positions, HSs' and L2s' H1-H2 values drastically dropped from the entire IP to IP-final position, mainly due to female HSs and female L2s.

Although no significant main effect was found for gender, significant interaction between position and gender ($\beta = 5.95$, $SE = 0.404$, $t = 14.736$) indicates that the H1-H2 difference between the two positions was larger for female speakers (baseline condition) than for male speakers. Moreover, there was a significant interaction between group and gender for the HSs ($\beta = -3.426$, $SE = 0.465$, $t = -7.364$). That is, the difference between female NSs' and female HSs' H1-H2 values was larger than the difference between those of male NSs and male HSs. Although it did not reach significance level, a similar tendency was found for the L2s. Pairwise comparisons of group and gender showed that while no significant group difference was found among the male speakers, both female HSs and female L2s had significantly

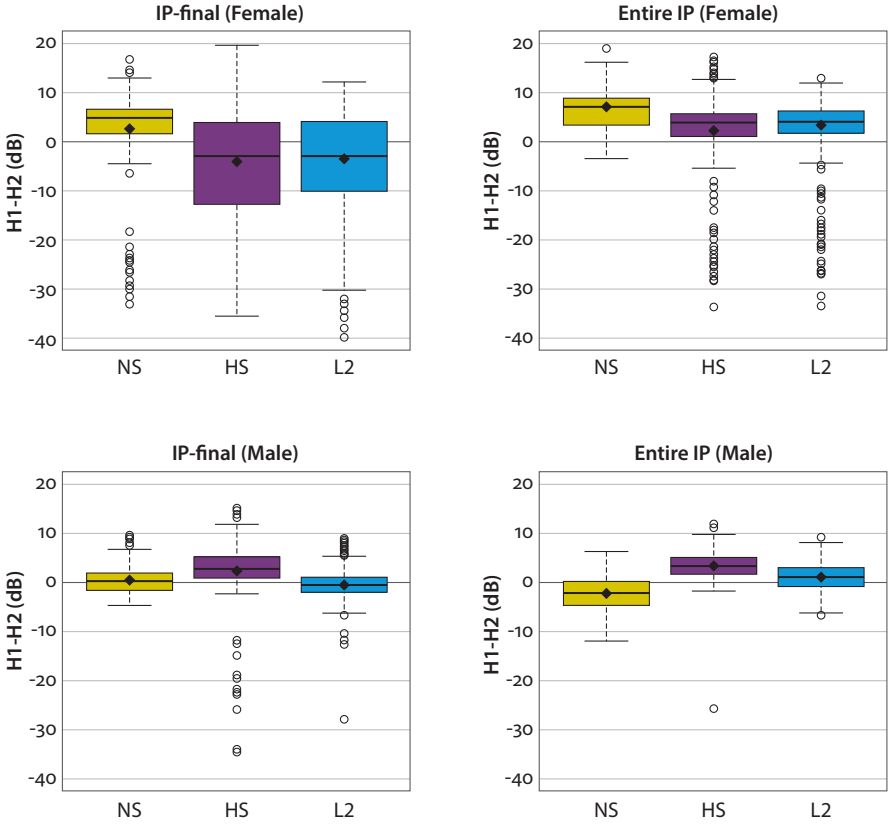


Figure 2. Relative amplitude of the first two harmonics (H1-H2) of the entire IP and in IP-final position of Spanish native monolinguals (NS), Spanish heritage speakers (HS), and English L2 learners of Spanish (L2), both male and female speakers (mean values are marked with filled diamonds)

lower H1-H2 values that female NSs (female HS: $p < 0.01$, female L2: $p < 0.05$). As seen in Figure 2, female HSs' mean H1-H2 value (−1.42 dB) was lower than that of female L2s (−0.27 dB), when compared to female NSs (4.49 dB), while male HSs had higher mean H1-H2 value (2.87 dB) than that of male L2s (0.57 dB), when compared to male NSs (−0.98 dB). This may have resulted in HSs' significantly larger difference between male and female H1-H2 values when compared to the NSs.

No significant three-way interaction was found among the three fixed factors (i.e., group, gender, and position), indicating that the two-way interactions provided sufficient information in interpreting the data. Indeed, pairwise comparisons of all the three factors confirmed that when comparing the H1-H2 values in the two positions across groups within the same gender, female NSs had significantly higher

H1-H2 values than female HSs ($p < 0.001$) and female L2s ($p < 0.01$) only in IP-final position, while the three male groups did not differ from each other in any of the two positions. Interestingly, when comparing the H1-H2 values of the two positions across genders within the same group, female HSs had significantly lower H1-H2 values than male HSs ($p < 0.05$) only in IP-final position. This can clearly be seen in Figure 2 in which more than half of the productions of female HSs' H1-H2 values in this position were negative, varying from 0 dB to as low as −35.81 dB, while the majority of male HSs' H1-H2 values were positive and did not vary much from 0 dB (i.e., with modal phonation). Similarly, female L2s' H1-H2 values in general showed a larger variation below 0 dB in IP-final position (from 0 dB to −30.91 dB, excluding 6 outliers), compared to the male counterparts (from zero to −6.35 dB, excluding 5 outliers). However, as the large overlap between the interquartile range of these two gender groups showed (female L2: from −9.91 dB to 4.01 dB, male L2: from −2.01 dB to 0.98 dB), the difference between the two groups did not reach significance level. Lastly, regarding the NSs, the pairwise comparisons showed that female NSs' H1-H2 values were significantly higher than male NSs ($p < 0.001$) for the entire IP. Given that 18% of female NSs' H1-H2 values, as opposed to none in the male NS data, were above 10 dB, it is likely that female NSs produced with a breathier phonation than male NSs, supporting the general trend that female voice tends to be breathier than male voice (Klatt & Klatt 1990).

When examining individual data, despite slight inter-speaker variability, participants' voice quality did not vary in a systematic manner regardless of the position, except for female HSs and female L2s. As seen in Figure 3 and 4, while the variation of female HSs' and female L2s' H1-H2 values increased drastically in the negative range in IP-final position, the rest of the participant groups maintained similar H1-H2 range across the two positions. With regard to the entire IP, although the majority of female HSs and female L2s produced within the modal phonation range, there were several outliers which were produced with low H1-H2 values (female HS: 33 out 430 cases, female NS: 28 out of 699 cases). Such instances rarely occurred in other participant groups. Since these outliers consisted of only 7% of the female HS data and 4% of the female L2 data, compared to 34.19% of the female HS data and 24.75% of the female L2 data in IP-final position, it is difficult to conclude that the creaky phonation found in IP-final position is simply due to their idiosyncratic speech style.

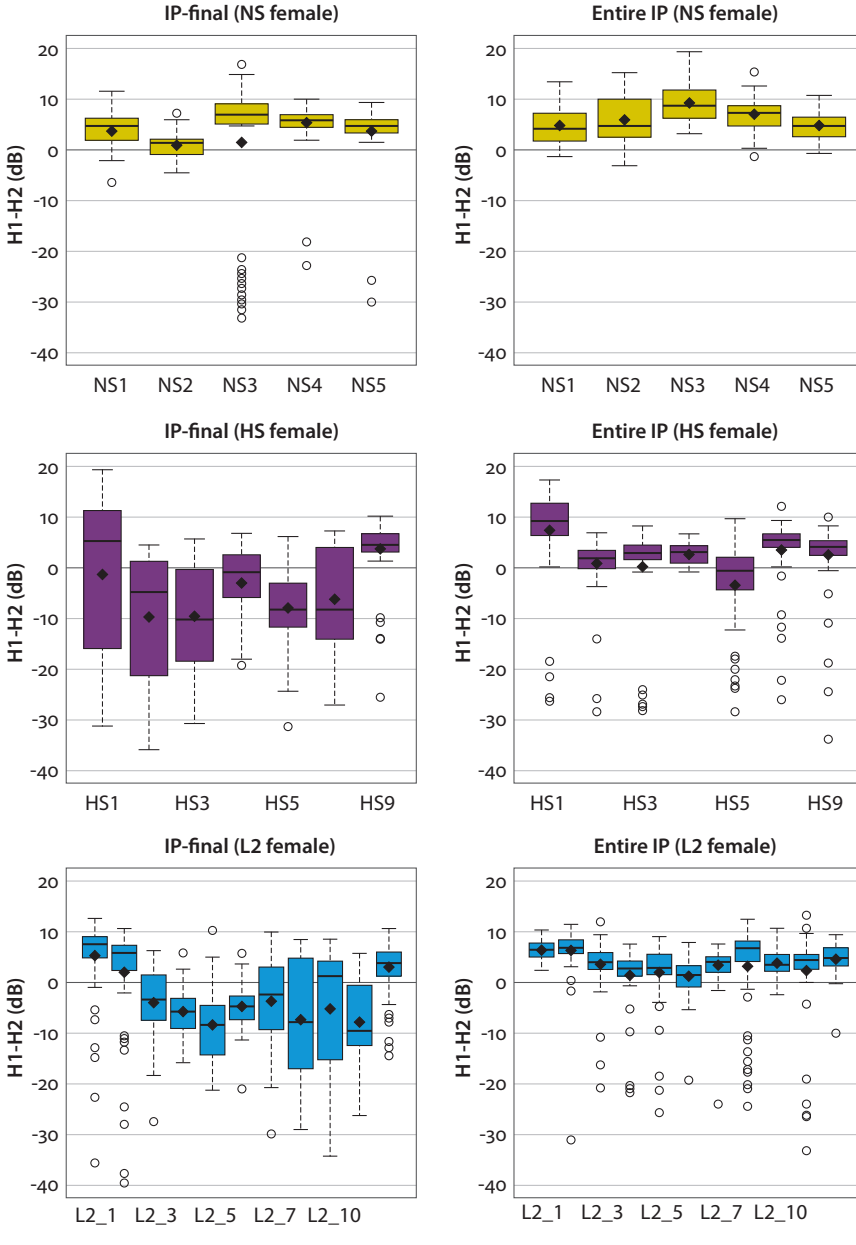


Figure 3. Individual female data (mean values are marked with filled diamonds)

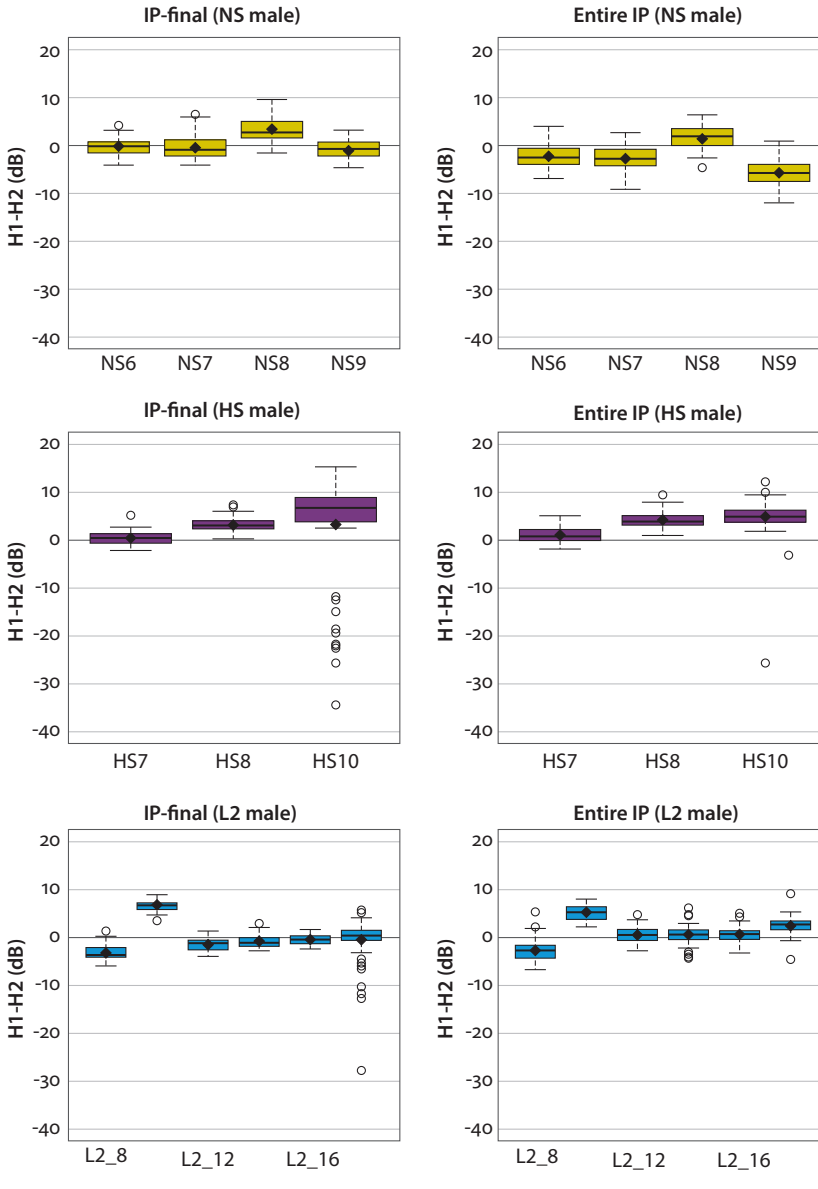


Figure 4. Individual male data (mean values are marked with filled diamonds)

Although creaky phonation has been found mostly in the speech of female HSs and female L2s, this is not to say that creaky voice is used only by these speakers, given that there were some exceptions in the present data. There was one female NS (NS3), one male HS (HS10), and one male L2 (L2_17) who had low H1-H2 values in IP-final position. However, given that these values only consisted of a minority

of their data and no systematic pattern was found among them, it is difficult to conclude that these speakers produced creaky phonation to the same extent as most of the female HSs and female L2s. With respect to female HSs and female L2s, there were one female HS (HS9) and three female L2s (L2_1, L2_2, L2_15) who maintained modal voice quality even in IP-final position, although more negative H1-H2 values were found in this position, compared to the entire IP.

5. Discussion and conclusion

The findings of the present study support that voice quality can be transferred from one language to another, even if it is not used contrastively in any of the two languages. Both groups of English-dominant bilinguals that participated in the study (i.e., HSs and L2s) used creaky voice in IP-final position when they produced sentences in Spanish and this pattern was observed mainly in the speech of female speakers. Given that creaky voice in IP-final position is a feature found in American English that indexes upwardly-mobile young American female persona (Podesva 2013; Todaka 1994; Wolk et al. 2012; Yuasa 2010), while it does not have such a socio-indexical meaning in Mexican Spanish (and most likely in other Spanish dialects as well), the use of creaky voice found in female HSs and female L2s can be considered as a transfer phenomenon of voice quality from American English to Spanish. It is interesting to note that creaky voice was equally found in both female HSs' and female L2s' speech in IP-final position, regardless of the differences in their linguistic and cultural background. The female HSs learned Spanish at home from birth and have stronger ties to Mexican culture, while the female L2s learned it as a second/foreign language and have minimal to no ties to Mexican culture. However, given that both groups are more dominant in English, it is reasonable to argue that language dominance has an effect in the transfer of creaky voice, similar to other phonetic features such as consonants and vowels (Flege et al. 1995; Flege et al. 1999; Grosjean 1989; Kuhl et al. 2003; Major 1992).

Despite such findings, there are several limitations in the present study that need to be addressed in future research. To begin with, the materials used in the present study were limited to one sentence type (i.e., declarative sentences with subject-verb inversion). However, this by no means excludes the possibility that voice quality may be transferred to varying degrees in different sentence structures. Therefore, future research needs to consider various sentence structures to have a fuller understanding of bilinguals' use of creaky voice. Moreover, although the findings showed a clear tendency that creaky voice is mainly used among female HSs and female L2s, based on the present data, in which the male data are underrepresented compared to the female one, it is difficult to conclude whether

creaky voice is only used by female HSs and female L2s, and never used by the male counterparts. As creaky voice is usually accompanied with low pitch (Johnson 2003), it may be observed in some male HSs and male L2s, and, if not, it would be interesting to examine whether male speakers refrain themselves from using this voice register to avoid being associated with the socio-indexical meaning attached to it. Furthermore, as the results showed that 1 out of 9 female HSs and 3 out of 11 female L2s did not produce creaky voice, it is likely that some speakers either are able to keep the voice quality of the two languages separate or do not use creaky voice in general to avoid the socio-indexical meaning with which it is associated. These exceptions suggest that considering bilinguals' language dominance alone may not be sufficient to explain their use of creaky voice in Spanish. Social factors such as cultural dominance may be an important factor that can explain the individual variability found in the present study, given that dominance is not limited to languages, but it can also be applied to bilinguals' two cultures. Grosjean (2010) argued that one culture often plays a larger role than the other in a bilingual speaker; thus, a bilingual speaker can be more dominant in one culture than the other, the same way as she/he can be more dominant in one language than the other. Therefore, future research on the degree of cultural dominance, such as the Bicultural Identity Integration (Benet-Martínez et al. 2002), is needed to examine the effect of cultural dominance on the transfer of creaky voice. Considering social factors will also provide an insight in whether not only the creaky voice *per se*, but the socio-indexical meaning associated with it (i.e., upwardly-mobile young American female persona) is also transferred from American English to Spanish.

References

Andrade, Argelia. 2003. "Intonational phonology of alteño Spanish." M.A. Thesis, University of California, Los Angeles.

Anderson, Rindy C., Casey A. Klofstad, William J. Mayew, and Mohan Venkatachalam. 2014. "Vocal Fry May Undermine the Success of Young Women in the Labor Market." *PLoS ONE* 9(5): e97506. doi:10.1371/journal.pone.0097506

Armstrong, Meghan E., Nicholas Henriksen, and Christian DiCanio. 2015. "Sociophonetic Analysis of Young Peninsular Spanish Women's Voice Quality." In *Hispanic Linguistics at the crossroads: Theoretical linguistics, language acquisition and language contact. Proceedings of the Hispanic Linguistics Symposium* 2013, ed. by Rachel Klassen, Juana M. Liceras, and Elena Valenzuela, 293–312. Amsterdam: John Benjamins Publishing Company. doi:10.1075/ihll.4.15arm

Baayen, R. Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: UK: Cambridge University Press. doi:10.1017/CBO9780511801686

Benet-Martínez, Verónica, Janxin Leu, Fiona Lee, and Michael W. Morris. 2002. "Negotiating biculturalism: Cultural frame-switching in biculturals with oppositional vs. compatible cultural identities." *Journal of Cross-Cultural Psychology* 33: 492–516. doi: 10.1177/0022022102033005005

Boas, Franz. 1947. "Kwakiutl Grammar with a Glossary of the Suffixes" *Transactions of the American Philosophical Society*, ed. by Helene Boas Yampolsky and Zelig Harris, 37(3): 201–377. doi: 10.2307/1005538

Brown, Penelope, and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Studies in Interactional Sociolinguistics 4. Cambridge: Cambridge Univ. Press.

Bruyninckx, Marielle, Bernard Harmegnies, Joaquim Llisterri, and Dolors Poch. 1994. "Language-induced voice quality variability in bilinguals." *Journal of Phonetics* 22: 19–31.

Coates, Jennifer. 1986. *Women, men, and language: A sociolinguistic account of sex differences in language*. London: Longman.

Colton Raymond H, Janina K. Casper, and Rebecca Leonard. 2011. *Understanding Voice Problems: A Physiological Perceptive for Diagnosis and Treatment*. Baltimore, MD: Lippincott Williams & Wilkins.

Dilley, Laura, Stefanie Shattuck-Hufnagel, and Mari Ostendorf. 1996. "Glottalization of word-initial vowels as a function of prosodic structure." *Journal of Phonetics* 24: 423–444. doi: 10.1006/jpho.1996.0023

Esling, John H. 1978. *Voice Quality in Edinburgh: A Sociolinguistic and Phonetic Study*, Ann Arbor: University Microfilms.

Esling, John H. 1982. "Pronunciation Consideration in ESL: Voice Quality Settings." *Working Papers of the Linguistics Circle of the University of Victoria* 2(1): 1–16.

Esling, John H. 1984. "Laryngographic study of phonation type and laryngeal configuration." *Journal of the International Phonetic Association* 14: 56–73. doi: 10.1017/S0025100300002796

Esling, John H., and Rita F. Wong. 1983. "Voice quality settings and the teaching of pronunciation." *TESOL Quarterly* 17(1): 89–95. doi: 10.2307/3586426

Esposito, Christina M. 2006. "The Effects of Linguistic Experience on the Perception of Phonation." Ph.D. dissertation, University of California, Los Angeles. 2006.

Esposito, Christina M. 2010. "The Effects of Linguistic Experience on the Perception of Phonation." *Journal of Phonetics* 38: 306–316. doi: 10.1016/j.wocn.2010.02.002

Flege, James E. 1992. "The intelligibility of English vowels spoken by British and Dutch talkers." In *Intelligibility in Speech Disorders: Theory, Measurement, and Management, Studies in Speech Pathology and Clinical Linguistics*, ed. by Raymond D. Kent, 157–232. Amsterdam, The Netherlands: John Benjamins Publishing. doi: 10.1075/sspcl.1.06fle

Flege, James E., Murray J. Munro, and Inn R. A. MacKay. 1995. "The effect of age of second language learning on the production of English consonants." *Speech Communication* 16: 1–26. doi: 10.1016/0167-6393(94)00044-B

Flege, James E., Grace H. Yeni-Komshian, and Serena Liu. 1999. "Age constraints on second language learning." *Journal of Memory and Language* 41: 78–104. doi: 10.1006/jmla.1999.2638

Fought, Carmen. 2003. *Chicano English in context*. New York: Palgrave Macmillan.

Gick, Bryan, Ian Wilson, and Donald Derrick. 2013. *Articulatory Phonetics*. Malden, MA: Wiley-Blackwell.

Golla, Victor K. 1970. "Hupa grammar." Ph.D. dissertation, University of California, Berkeley. 1970.

Gordon, Matthew, and Peter Ladefoged. 2001. "Phonation Types: A Cross-linguistic Overview." *Journal of Phonetics* 29: 383–406. doi: 10.1006/jpho.2001.0147

Grivicic, Tamara, and Chad Nilep. 2004. "When Phonation Matters: The Use and Function of *yeah* and Creaky Voice." *Colorado Research in Linguistics* 17: 1–11.

Grosjean, François. 1989. "Neurolinguists, beware! The bilingual is not two monolinguals in one person." *Brain and Language* 36: 3–15. doi: 10.1016/0093-934X(89)90048-5

Grosjean, François. 2010. *Bilingual: Life and reality*. Cambridge, Mass: Harvard University Press. doi: 10.4159/9780674056459

Harmegnies, Bernard, and Albert Landercy. 1985. "Language Features in the Long-term Average Spectrum." *Revue de Phonétique Appliquée* 73–75: 69–80.

Harmegnies, Bernard, Marielle Bruyninckx, Joaquim Llisterri J, and Dolors Poch. 1989. "Effects of Language Change on Voice Quality. An Experimental Contribution to the Study of Catalan-Castillian Case." *Proceedings of the 1st European Conference on Speech Communication and Technology*, 489–492.

Henton, Caroline, and Anthony Bladon. 1988. "Creak as a Sociophonetic Marker." *Language, Speech and Mind: Studies in Honour of Victoria A. Fromkin*, ed. by Larry M. Hyman and Charles N. Li, 3–29. London: Routledge.

Heselwood, Barry, and Zahra Mahmoodzade. 2007. "Vowel Onset Characteristics as a Function of Voice and Manner Contrast in Persian Coronal Stops." *Leeds Working Papers in Linguistics and Phonetics*, ed. by Barry Heselwood and Cécil De Cat, 12: 125–142.

Johnson, Keith. 2003. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell.

Keating, Patricia A., and Christina Esposito. 2006. "Linguistic Voice Quality." *UCLA Working Papers in Phonetics* 105: 85–91.

Kim, Ji-Young. "The Use of Creaky Voice by Spanish Heritage Speakers in the US." Poster Presentation at the 7th Heritage Language Research Institute, Chicago, IL, June 17–21, 2013.

Kirk, Paul L., Jenny Ladefoged, and Peter Ladefoged. 1993. "Quantifying Acoustic Properties of Modal, Breathy and Creaky Vowels in Jalapa Mazatec. *American Indian Linguistics and Ethnography in Honor of Laurence C. Thompson*, ed. by Anthony Mattina and Timothy Montler, 435–450.Missoula, MT: University of Montana Press.

Klatt, Dennis H., and Laura C. Klatt. 1990. "Analysis, Synthesis, and Perception of Voice Quality Variations among Female and Male Talkers." *Journal of the Acoustical Society of America* 87: 820–857. doi: 10.1121/1.398894

Kreiman, Jody, Bruce R. Gerratt, and Sameer U. Khan. 2010. "Effects of Native Language on Perception of Voice Quality." *Journal of Phonetics* 38: 558–593. doi: 10.1016/j.wocn.2010.08.004

Kuhl, Patricia K., Feng-Ming Tsao, and Huei-Mei Liu. 2003. "Foreign language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning." *Proceedings of the National Academy of Sciences* 100: 9096–9101. doi: 10.1073/pnas.1532872100

Ladefoged, Peter. 1971. *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago.

Ladefoged, Peter. 1993. *A Course in Phonetics*. Fort Worth, TX: Harcourt Brace College Publishers.

Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.

Laver, John. 1980. *The Phonetic Descriptions of Voice Quality*. Cambridge: Cambridge Univ. Press.

Major, Roy C. 1992. "Losing English as a first language." *The Modern Language Journal* 76: 190–208. doi: 10.1111/j.1540-4781.1992.tb01100.x

Mendoza, Elvira, Nieves Valencia, Juana Muñoz, and Humberto Trujillo. 1996. "Differences in Voice Quality Between Men and Women: Use of the Long-Term Average Spectrum (LTAS)." *Journal of Voice* 10 (1): 59–66. doi: 10.1016/S0892-1997(96)80019-1

Mendoza-Denton, Norma. 2011. "The semiotic hitchhiker's guide to creaky voice: Circulation and gendered hardcore in a Chicana/o gang persona." *Journal of Linguistic Anthropology* 21: 261–280. doi: 10.1111/j.1548-1395.2011.01110.x

Morrison, Geoffrey S., and Paola Escudero. 2007. "A Cross-dialect Comparison of Peninsular- and Peruvian-Spanish Vowels." *Proceedings from the XVI International Congress of Phonetic Sciences*, ed.by Jürgen Trouvain, and William J. Barry, 1505–1508. Saarbrücken: Germany.

Nichols, Jenna. "Indexicality and Voice Quality in a French-English Bilingual." Paper Presentation at the UIC Bilingualism Forum, Chicago, IL, October 4–5, 2012.

Ohala, John. 1994. "The frequency code underlies the sound-symbolic use of voice pitch." *Sound Symbolism*, ed. by Leanne H. Hinton, Johanna Nichols, and John Ohala, 325–347. Cambridge: Cambridge University Press.

Pittam, Jeffery. 1987. "Listeners' evaluation of voice quality in Australian English speakers." *Language and Speech* 30: 99–113.

Podesva, Robert J. 2007. "Phonation type as a stylistic variable: The use of falsetto in constructing a persona". *Journal of Sociolinguistics* 11(4): 478–504. doi: 10.1111/j.1467-9841.2007.00334.x

Podesva, Robert J. 2013. "Gender and social meaning of non-modal phonation types." *Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society* ed. by Chundra Cathcart, I-Hsuan Chen, Greg Finley, Shinae Kang, Clare S. Sandy, and Elise Stickles, 427–448.

Quenqua, Douglas. "They're, Like, Way Ahead of the Linguistic Currrrve". *The New York Times*. Last modified February 27, 2012. http://www.nytimes.com/2012/02/28/science/young-women-often-trendsetters-in-vocal-patterns.html?_r=4&pagewanted=all%3Fsrc%3Dtp&smid=fb-share&

Stuart-Smith, Jane. 1999. "Glasgow: Accent and voice quality." *Urban Voices* ed. by Paul Foulkes and Gerard J. Docherty, 203–222. London: Arnold.

Tajfel, Henri. 1974. "Social identity and intergroup behavior." *Social Science Information* 13: 65–93. doi: 10.1177/053901847401300204

Todaka, Yuichi. 1994. "A Cross-language Study of Voice Quality: Bilingual Japanese and American English Speakers." Ph.D. dissertation, University of California, Los Angeles. 1994.

Trittin, Pamela J., and Andrés de Santos y Lleó. 1995. "Voice quality analysis of male and female Spanish speakers." *Speech Communication* 16: 359–368. doi: 10.1016/0167-6393(95)00004-8

Wolk, Lesley, Nassima B. Abdelli-Beruh, and Dianne Slavin. 2012. "Habitual use of vocal fry in young adult female speakers." *Journal of Voice* 26: e111–e116. doi: 10.1016/j.jvoice.2011.04.007

Yuasa, Ikuko P. 2010. "Creaky voice: A new feminine voice quality for young urban-oriented upwardly mobile American women?" *American Speech* 85: 315–337. doi: 10.1215/00031283-2010-018

CHAPTER 10

Null subjects in the early acquisition of English by child heritage speakers of Spanish

Jennifer Austin, Liliana Sánchez and Silvia Perez-Cortes
Rutgers University

This paper investigates the development of English and Spanish subject production in 8 heritage bilingual children (age range: 4;1–5;3; mean age: 4;7). Oral production of overt and null subjects was assessed using a picture-based story re-telling task and a description task. Subjects were coded according to type (overt or null), and pragmatic function (new information, topic continuation, recovery, contrast and change of topic). Results indicate a robust distribution of pragmatically appropriate subjects in Spanish; however, non adult-like null subjects were also found in a subset of the English utterances. We propose that cross-linguistic influence from Spanish encouraged children to have more than one grammar available (Amaral & Roeper 2014), with the option to treat English as a null-subject language.

Keywords: heritage bilinguals, child bilingualism, cross-linguistic influence, null subjects, acquisition of Spanish, acquisition of English

1. Introduction

Young bilingual children who are dominant in an overt-subject language have been found to produce pragmatically infelicitous overt subjects in their agreement-based null-subject language (Paradis & Navarro 2003; Serratrice et al. 2004; Hacohen & Schaeffer 2007, a.o.). These observations are consistent with proposals claiming that cross-linguistic influence (CLI) is more likely to occur at the syntax-pragmatics interface than in other areas of grammar (Hulk & Müller 2000), although recent findings have refined this claim, extending the effects of CLI to non-interface phenomena (Cuza 2012; Slabakova & Ivanov 2011), underscoring the role of language dominance (Liceras, Fernández Fuertes, de la Fuente, Boudreau & Acevedo 2012b; Yip & Matthews 2000) and the potential effects of cross linguistic priming (Pérez-Leroux, Cuza & Thomas 2011). The production of infelicitous overt subjects in

DOI 10.1075/rllt.11.10aus

Spanish has also been found among adult Spanish heritage speakers (Montrul 2004, 2009), who are typically defined as early Spanish-English bilinguals living in an English-dominant society who become gradually more dominant in the majority language as they enter the school system (Pires & Rothman 2009). The higher frequency of overt pronominal subjects in the Spanish of adult English-dominant heritage speakers, as seen in (1), has been attributed to their incomplete acquisition of the pragmatic constraints governing the distribution of null and overt subjects in Spanish (Montrul 2004: 133):

(1) Había una vez una niña chiquita que se llamaba Caperucita. Ella vivía con su mamá y *ella quería mucho a su abuelita.
'Once upon a time there was a little girl whose name was Little Riding Hood. She lived with her mother and she loved her grandmother very much'

Interestingly, and as noted by Liceras et al.'s (2012) studies, Spanish heritage children who receive a more balanced input do not seem to produce higher rates of overt subjects when compared to Spanish-dominant controls.

In contrast with some of the previous findings, monolingual children acquiring languages which require overt subjects have been reported to go through a period of subject omission in obligatory contexts until approximately age 3;0 (Bloom 1970; Hyams 1986; Hyams & Wexler 1993; Valian 1990). The adoption of a temporary pro-drop option in these languages is linked to the gradual acquisition of the inflectional system (Pierce 1992) and has been observed to co-occur with root infinitives (Sano & Hyams 2000) and copula omission. There is debate in the field as to whether these null-subject productions reflect a grammar that is syntactically divergent from adult-like representations (Orfitelli & Hyams 2012) or whether they are attributable to processing problems (Bloom 1990). These developmental patterns raise the question of how crosslinguistic influence (CLI) affects the licensing of subjects in the early acquisition of English by bilingual children who are heritage speakers of Spanish, given that child L1 English allows for null subjects and bare verbs.

The purpose of this preliminary inquiry is to investigate the development of subject production in English and Spanish in 8 heritage bilingual children (ages 4;1–5;3), who have acquired a non-Caribbean variety of Spanish at home and are experiencing an increase in exposure to English as a language of instruction at school. The age range was chosen in order to compare their data to those of English and Spanish monolingual children. At this age range, Spanish monolingual children have already acquired all the properties associated with null subjects (Austin, Blume, Parkinson, Proman, Lust & Núñez del Prado 1997), and English monolingual children have shifted to a more consistent production of overt subjects (Hyams 1986). If exposure to Spanish has been consistent, one would expect these bilingual children to have acquired null subjects in Spanish. At the same time, these children

are at the early stages of a major shift towards greater exposure to English through schooling. In contrast to previous research on bilingual Spanish-English children, which have focused on the distribution of null subjects in either Spanish or English (e.g. Paradis & Navarro 2003), we looked for evidence of CLI in the production of null subjects in both of the children's languages at the early stages of the shift in exposure to English. We also look at the pragmatic functions of null and overt subjects in both languages.

2. Analyzing null and overt subjects

2.1 Verbal morphology and the distinction between null and overt subject languages

English has been considered an overt-subject language since the early proposals that distinguished between languages with overt subjects and languages with null subjects (Chomsky & Lasnik 1977; Chomsky 1981). The obligatory nature of overt subjects in languages like English was derived from the Extended Projection Principle (EPP), according to which all sentences must have a syntactic subject (Chomsky 1981). In non-null subject languages (i.e. English), this principle must be satisfied by overt pronouns or determiner phrases (DPs). Null subject languages satisfy the EPP by means of verbal agreement morphology that allows for the licensing (a condition on the syntactic representation) and identification (a condition on the interpretation) of a null pronoun as a subject as illustrated by this contrast:

(2) $(pro)_i$ bail-a_i
$(pro)_i$ dance-3s
'(He/she/it) dances.'

(3) He/She/It dances

Later reformulations of the EPP have proposed to reinterpret it as an abstract feature that must be valued (Alexiadou & Anagnostopoulou 1998; Chomsky 1995; Lasnik 2001; Picallo 1998). Chomsky (1995) operationalizes the EPP as an abstract feature (+D) carried by Tense that must be checked either by agreement morphology on the verb or by a DP in the specifier of Tense Phrase (TP). Overt-subject languages like English lack the appropriate type of verbal agreement morphology with pronominal features to satisfy the EPP requirement. In their absence, a full DP projection is attracted to the specifier of TP to check the D feature in Tense (T).[1]

1. For a detailed review of the proposals concerning this distinction, see Camacho (2013).

We will assume that the main difference between English and Spanish is the non-pronominal vs. pronominal nature of their respective verbal morphology. While the Spanish verbal morpheme *-a* in (2) is pronominal in nature and, therefore, able to satisfy the D feature in T-, the English morpheme -e*s* in (3) is not.

More specifically, Camacho (2013) proposes that there is parametric variation across languages on the conditions that verbal agreement must meet in order to identify null subjects. These stipulations are articulated in his Minimal Morphological Threshold (MMT): "The minimal morphological threshold (MMT) defines the minimal set of values overtly encoded in the morphology that a language requires to identify a null subject" (Camacho 2013: 114). Languages such as Spanish and English differ crucially with respect to the MMT. While a language like Spanish is able to identify null subjects by recovering the person and number values encoded in its verbal morphology, as in (2), a language like English, which lacks a rich verbal paradigm, must resort to a full pronoun instead, as illustrated in (3).

2.2 Pragmatic conditions on null and overt subjects in Spanish

In addition to the differences in the syntactic conditions for licensing and identification of subjects, Spanish and English also differ with respect to the relationship between subjects and discourse topics. Overt subjects in Spanish have traditionally been analyzed as constituents in A' positions (Contreras 1991; Ordoñez & Treviño 1999; Zubizarreta 1999; but see Suñer 2003 for an alternative analysis). In most proposals, unstressed overt subjects in Spanish occupy a topic position. Overt subjects in English, on the other hand, are typically assumed to be in Spec of TP (Chomsky 1995).

While null subjects in Spanish are preferred when they refer to a previously introduced topic in discourse (Frascarelli 2007), overt subjects in Romance languages correlate with the need to introduce a new topic (Alonso-Ovalle, Fernández-Solera, Frazier, & Clifton 2002). This preference is formalized in Frascarelli (2007) as a Topic Criterion according to which an Aboutness Topic, an extension of the EPP feature that serves the discourse function of introducing a new topic, is projected in the Complementizer Phrase (CP) area. The EPP feature in the Aboutness Topic projection matches with an argument in the main clause through the syntactic operation Agree. This matching allows the sentence level topic and the subsequent subjects with the same referent to be null throughout discourse. We will adopt Frascarelli's view according to which the grammar of Spanish allows for the continuous licensing of null subjects in discourse related to the same Aboutness Topic. In our view, acquisition of the distribution of null and overt subjects in Spanish

has two components: the acquisition of verbal morphology and the acquisition of the Topic Criterion.

3. Early acquisition of null and overt subjects

Over the last two decades, there has been a considerable amount of research dedicated to the acquisition of subjects in null and overt subject languages (Austin et al. 1996, 1997; Bel 2003; Bloom 1990; Grinstead 2004; Hyams 2011; Hyams & Wexler 1993; Orfitelli & Hyams 2012; Villa-Garcia 2012). In this section, we summarize the most relevant studies regarding the acquisition of null and overt subjects by English and Spanish monolingual children, heritage bilinguals, and L2 children.

Research conducted with young monolingual children (ages 2;0–3;0) of overt-subject languages such as English or German have reported a consistent pattern of null subject use in contexts where an overt subject would be expected (Austin et al. 1996; Bloom 1990; Hyams 2011; Valian 1991). Although the frequency of subject omission seems to vary across children (average production in spontaneous speech: 10%–57%), it is believed to converge with adult-like patterns at around age 3;5 (Orfitelli & Hyams 2012). This null-subject stage has been traditionally explained by two different types of accounts: those favoring processing difficulties (Bloom 1990; Valian 1991), and those alluding to a generalized (but temporary) adoption of a pro-drop grammar (Hyams 1994; Hyams & Wexler 1993). In what follows, we will review some key observations regarding this last account given its potential implications for the present study.

Some researchers have argued that English-speaking children show the same pragmatic distribution as Spanish or Italian-speaking children when licensing null subjects (Guerriero, Cooper, Oshima-Takane & Kuriyama 2001; Hughes & Allen 2006). However, in contrast with languages like Spanish, Italian or Chinese, subject omission in child English is often associated with the presence of underspecified inflection, which would license a PRO in subject position (Sano & Hyams 2000).

Previous studies on the acquisition of subjects by simultaneous bilingual children[2] have provided strong evidence for CLI (Argyri & Sorace 2007; Paradis & Navarro 2003; Serratrice 2006). Specifically, it has been proposed that the language with the least economical system, that is "(the one) with both null and overt pronouns, where the later are associated with focus and topic shift" will be more likely

2. We will follow Blom and Unsworth's (2010) criteria to differentiate between simultaneous and sequential bilinguals. While the former have been regularly exposed to dual language learning since birth (or before age 3;0), sequential bilinguals "have had significantly less exposure to their L2 than their L1" (237).

to be influenced by the most economical system (Serratrice 2006: 191). This has been shown to be the case of children learning null-subject languages like Hebrew, Italian or Spanish at the same time as English, an overt-subject language (Hacohen & Schaeffer 2007; Paradis & Navarro 2003; Serratrice, Sorace, & Paoli 2004; Sorace, Serratrice, Filiaci & Baldo 2009). While there seems to be a widespread tendency to overproduce overt subjects in bilinguals' null-subject languages between the ages of 1;09–4;06, work with older children (between ages 6;0–10;0) showed that this pattern only continues if children are dominant in the overt-subject language (Serratrice 2006; Sorace & Serratrice 2009).[3] According to Liceras et al. (2012b) and Sorace et al. (2009), the opposite trend, that is, to transfer null subjects from a null-subject language to an overt one, is highly unlikely. Liceras et al. (2012a) hypothesized that the lack of robust evidence for null subjects in English, as well as the impossibility of combining them with inflected verbs would prevent children from transferring the distribution of null and overt subjects from Spanish to English, thus converging into a pro-drop grammar. Liceras et al.'s (2012b) most recent study, however, shows that in the case of Spanish-English bilinguals living in Spain, language dominance in the null-subject language could play a role in the development of English. Production data from 2 participants showed that Spanish, which they characterized as having lexical specialization, namely, strong pronouns have pragmatic value while subject agreement morphemes only have syntactic value, seemed to influence English, a language with one set of pronouns with pragmatic and syntactic value, by reducing the rate of null subjects and root infinitives in comparison to age-matched English monolinguals. In their view, the fact that the Spanish pronominal system is more complex has an accelerating effect on the acquisition of the simpler English pronominal system.

Language dominance also seemed to play a role in the sequential Spanish-English bilinguals analyzed by Lakshmanan (1994), who were learning English as an L2 via immersion. These two L1 Spanish children aged 4;6 (Marta) and 5;0 (Cheo) had been immersed in English for less than three months before the beginning of the study, and were initially reported to produce a high rate of null subjects in English, which mostly co-occurred with inflected verbs (Hilles 1991; Lakshmanan 1994).[4] Marta's data show an initially high production of null subjects

3. Contrary to these studies, Hinzelin (2003), Juan-Garau & Pérez-Vidal (2000) and Liceras et al. (2012a, 2012b) report that young bilinguals do not seem to necessarily exhibit difficulty in "setting the null subject parameter to its respective value in each of their two languages" (Hinzelin 2003: 19).

4. Martha was born in Puerto Rico. Prior to her arrival in the U.S., she attended a nursery school where Spanish was the language of instruction. Cheo was born in Colombia and had no prior exposure to English before arriving in the U.S. (Lakshmanan 1994: 74).

in English (around 20% in months 2–4) and a high production of inflected verbs during the same time (around 75%). Cheo's production of null subjects for the first few months of the study averaged around 10%, and his production of inflected verbs averaged 60% (with considerable fluctuations between the third of fifth month of recordings). Unlike the null subjects reported in monolingual English children, which have often been associated with the lack of verbal inflection (Sano & Hyams 2000), both Marta and Cheo showed a preference for using null subjects with inflected verbs in English, mirroring their dominant language (Spanish). Our study, which also shows the same preference for Spanish heritage children in their first stages of instructed English immersion, complements Lakshmanan's by providing possible evidence of CLI from the children's dominant language. Although limited, the data obtained from 8 Spanish heritage children shows that while they seemed to have a robust distribution of pragmatically appropriate null and overt subjects in Spanish, they exhibit non-target null subjects in their least dominant language (English) following Spanish pragmatic constraints. Additionally, the present study was designed to extend their findings to older heritage populations (mean age 4;7) who had recently begun their schooling in the majority language (English).

3.1 Research questions

In light of these previous findings, we explore the following research questions:

1. To which degree do Spanish-English child heritage bilinguals show evidence of knowledge of the licensing and identification properties of null and overt subjects in Spanish and English?
2. In a context of a shift towards increased exposure to the socially dominant language (English), does prior exposure to Spanish play a role in their distribution of null and overt subjects in English? If so, how?

We hypothesize that children's predominant exposure to and dominance in Spanish should allow them to properly identify and use the discourse-pragmatics behind null/overt subject use in this language at the early stages of the shift in exposure. Furthermore, following Hilles' (1991) and Lakshmanan's (1994) studies on sequential bilinguals, we hypothesize that this particular group of heritage bilinguals, who have been exposed to Spanish at home from birth and have had lower levels of exposure to English at home before entering the school are likely to extend Spanish's licensing properties to English null subjects, associating this type of subjects with the continuity of an Aboutness Topic. We expected this trend to be reinforced by the high percentage of Spanish-speakers in the community (83.7%) and the distinct preference for the use of Spanish within the family.

4. The present study

4.1 Participants

A total of eight ($N = 8$) subjects (age range: 4;1–5;3; mean age: 4;7) were interviewed for this exploratory study, obtaining a total of 438 verb tokens. All children attended a public pre-school located in a northeastern city in the US with a high index of dominant Spanish speakers. 83.7% of the city's population speaks Spanish at home (American Community Survey, 2013), a tendency closely followed by the student body of the school (62.3%). These percentages, higher than the state (15.4%) and the country average (12.9%), are very representative of the subjects' linguistic background. Parents completed a language history questionnaire about their children's age, place of birth, language dominance and perceived proficiency, as well as country of origin of the family and preferences of language use at home, school and socially. As a point of comparison, we also analyzed Spanish and English narratives from monolingual children within the same age range (4;0–5;0 year old), publicly available in the CHILDES database (MacWhinney 2000). The Spanish data was obtained from López-Ornat's et al. (1994) 10 recordings of María (age: 4;0; total number of tokens: 391), as well as from Sebastián's (1989) frog story datasets from 9 children (age range 4;0–4;10, average: 4;5), reaching a total of 322 tokens. The English control data was collected by Renner (1988) and Marchman (1989, later reported in Berman & Slobin 1994), and included 35 age-matched English monolingual children (age range: 3;0–5;0; mean age: 4;4), which rendered a total of 640 verbal tokens.

With the exception of 3 of our bilingual children, who had acquired both Spanish and English simultaneously, all the remaining participants had been formally exposed to English after age 3;0 (62.5%). Children's families originated from three different countries: Ecuador (50%), El Salvador (25%) and Honduras (25%), none of which exhibit any distinct variation of the property under study. All but three of the families responded that they were more comfortable using Spanish at home. Two of them responded that they used both languages equally, whereas one preferred the use of English. Parents conveyed that 50% of the children spoke Spanish *like natives*, 25% *very fluently* and the remaining 25% *with some difficulty*. With regards to their command of English, 25% of the children were rated at *like native* level, 25% at *very fluently*, 12.5% as having *some difficulty* and 37.5% as having *very little/no knowledge at all*.

Given the high percentage of Spanish speakers in the area, and given the data obtained in the parental language background questionnaires, we would like to

venture that despite being born in the US, these children are in the earlier stages of experiencing a shift towards increased exposure to English in comparison to the input received in that language before attending pre-school. This observation is supported by our assessment of their proficiency in each of the languages, analyzed by an independent measure of linguistic complexity (Park, Tsai, Liu & Lau 2012), which calculated the total number of verbs in each language and the average number of words per verb phrase used in the narrative (p. 230), as seen in Figure 1.

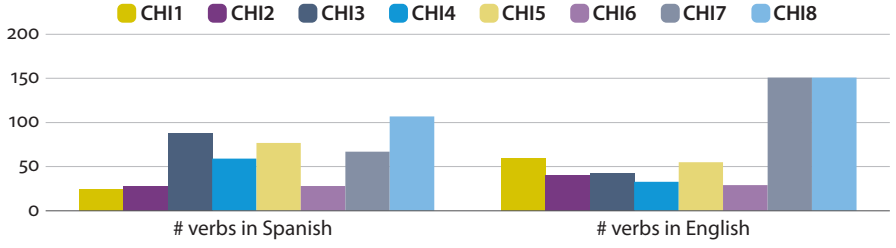


Figure 1. Mean verbs per narration as a function of discourse matrix language

Although initial results seemed to point to an overall dominance in English (mean verbs per narration in English (n = 70) was higher than the one observed in Spanish (n = 59.75)), a more detailed analysis pointed to a couple of outliers who reverted the trend towards a Spanish dominance across children (mean verbs per narration: 43 in English; 59.75 in Spanish), confirming the information obtained in the language history questionnaires.

4.2 Methodology

All bilingual children were interviewed during class time every six months during a two-year period, although this paper will only focus on data obtained in the first session for each language. Interviews lasted an average of 15 minutes, and in order to control for language mode, researchers collected data in each language at least one week apart. During the experimental session, participants were instructed to complete two production tasks: a story re-telling based on Mayer's (1969) *Frog Where are You?* series, and a description of images depicting different everyday situations (i.e. going to the dentist, attending a birthday party or celebrating Halloween). Data from both tasks were included in our analyses. All interviews were transcribed following CHAT conventions (MacWhinney 2000), and subsequently coded based on the system reproduced in the following table:

Table 1. Summary of the coding used in the present study

Verbal information	Subject realization	Other
– Argument structure – Verb type (auxiliary vs. main verb) – Location of the verb (main or subordinate clause) – Morphology (Tense, Aspect, Person, Number)	– Type of subject (overt DP/NP, overt pronoun, expletive pronoun, null subject, ambiguous) – Antecedent – Pragmatic function (new information, topic continuation, recovery, contrastive, change of topic)	– Presence of negation – Language of the utterance – Verboids (non-finite forms) – Errors

Although this paper will focus on information related to subject realization, one of the most relevant contributions of this study is the addition of more detailed categories in the coding for the discourse-pragmatic functions licensing the presence of null and overt subjects. The importance of these interface conditions in the prediction of potential locus of CLI led us to the incorporation of 5 new ways of classifying subjects based on their discourse function: (1) new information: there is no previous mention of the subject in question; (2) topic continuation: the subject is the same discourse topic; (3) topic retrieval: the subject is used to disambiguate between possible antecedents; (4) contrastive: the subject is opposed to a previous topic; and (5) change of topic: the subject reintroduces a previously mentioned theme and abandons the most recent topic. Data transcription and coding were completed independently by four researchers, and thoroughly discussed in order to ensure a high degree of inter-rater reliability.

5. Results

As shown in Figure 2, the Spanish data show evidence of a robust distribution of pragmatically-appropriate null and overt subjects, following the monolingual patterns observed in López-Ornat et al. (1994) and Sebastián (1989).

Despite the small size of the samples (713 verb tokens in the case of monolinguals, and 438 in the bilingual recordings) both datasets present fairly similar distributions of null and overt subjects based on specific pragmatic purposes. Additionally, both groups show similar overall percentages of null and overt subject use (monolinguals: 57.7% of overt subjects and 42.2% of nulls; bilinguals: 54.3% and 45.6% respectively). These results contrast with data from simultaneous bilinguals reporting CLI from English into Spanish at the pragmatics/syntax interface (Liceras et al. 2012b; Paradis & Navarro 2003).

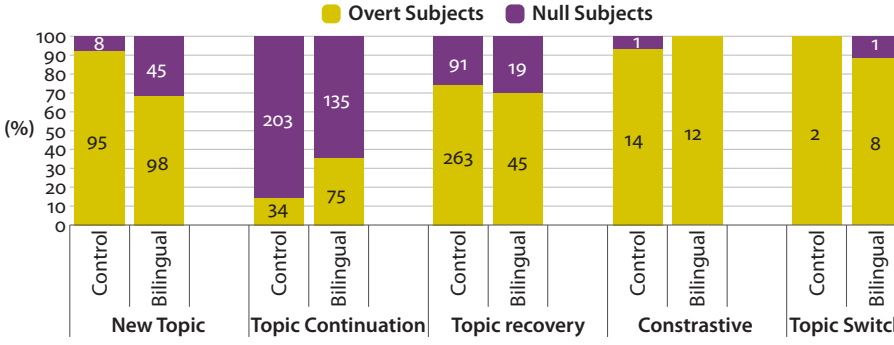


Figure 2. Distribution (%) of Spanish null/overt subjects in monolinguals and bilinguals according to discourse function (including the # of tokens in each category)

The only categories where bilingual children in our study seem to differ from Spanish monolinguals are those of new and continuous topic expression, which have been claimed to be the most vulnerable to CLI (Shin & Otheguy 2009). Although the tendency of both groups is ostensibly similar – preference for overt subjects to introduce new information and null subjects to express continuity –, it is predicted that more extensive contact with English is likely to affect these areas (Otheguy, Zentella & Livert 2007).

When compared with their English monolingual counterparts, the bilingual children interviewed show higher rates of null subjects, especially for a group of children that should be well beyond the "null-subject stage" (overall percentages per child range from 3–17%, average: 8.59% in bilinguals, but only 1–6.6%; average: 2.16% in age-matched monolinguals).

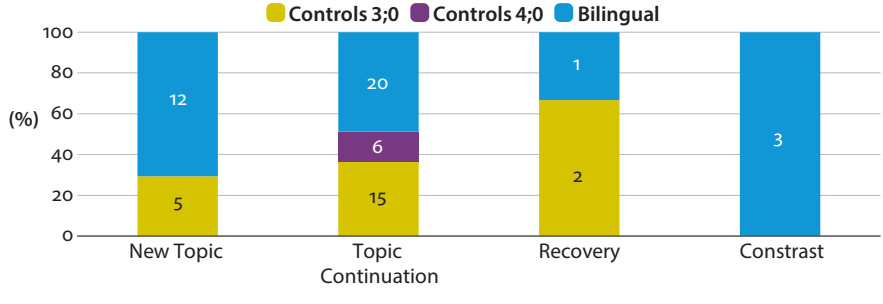


Figure 3. Distribution of English null subjects in monolinguals and bilinguals according to discourse function (# of tokens provided in the table below)

In addition, Figure 3 shows that both populations, monolingual and bilingual children, are most likely to produce null subjects in contexts of topic continuation,

such as (4), or when a new topic is introduced, as in the following example from a bilingual child:

(4) *EXP: yeah that's right and what was the boy doing?
*CHI: sand castle.
*EXP: yeah?
*CHI: yeah two of them.
*EXP: how do you know?
*CHI: because Ø have a bucket.

Figure 3 also indicates that there is more overlap between the contexts in which bilinguals and younger monolingual children produce null subjects (in new topic, topic continuation, and recovery environments) than between bilinguals and older monolingual children. However, unlike either group of monolingual children, bilingual children produced null subjects in contrastive contexts.

Previous studies had reported the co-occurrence of null subjects in English L1 acquisition with known and "accessible" referents (Guerriero et al. 2001; Hughes & Allen 2006; Hyams 2011), a pragmatic constraint that also operates in null-subject languages such as Spanish. The presence of null subjects in monolingual English acquisition, however, has also been traditionally associated with underspecified inflection (Rizzi 1993, 2005), as seen in Example (5) from our Spanish/English bilinguals:

(5) *EXP: and what did they do?
*CHI: Ø say froooooog!

However, as noted in Hyams (2011) there is still "a significant number of null subjects in (English) finite contexts" (p. 16), such as (6), which would point towards the initial adoption of a null subject language grammar by English monolingual children.

(6) *EXP: can you tell me what happened in this picture?
*CHI: the boy #the dog and the boy's carrying a fish and a bucket # and a [/] a turtle and the frog.
*EXP: mhm.
*CHI: Ø are running.

A more detailed analysis of the contexts in which subject omission took place in (Figure 4) seems to partially support Hyams' (2011) observations.

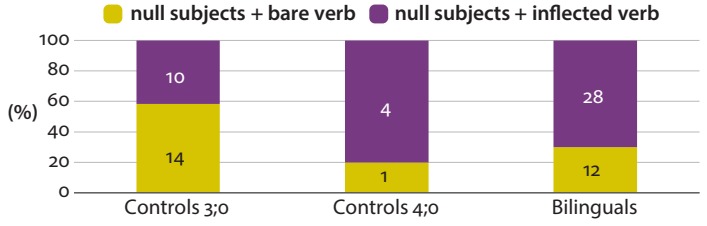


Figure 4. Rate of co-occurrence of null subjects with bare/inflected verbs (# of tokens provided in the table below)

In line with previous research (Bloom 1990; Hyams & Wexler 1993, *inter alia*) the spontaneous production of null subjects in monolingual English children seems to decrease over time (from 6.6% to 1.9%). While null subjects in younger children (mean age 3;5) appear to be more likely to co-occur with bare verbs (following Sano & Hyams 2000), they are visibly associated with the presence of verbal inflection in older children (mean age: 4;4) and bilinguals (mean age: 4;7), where the rates of co-occurrence of null subjects and inflected verbs oscillate between 68%–80%. A chi-square test showed a significant difference in the production of null subjects with inflected verbs between the three-year-old English monolinguals and the bilinguals ($2 = 4.99$, $df = 1$, $p < 0.05$). In contrast, there was no significant difference in the rate of production of null subjects with inflected verbs between the four-year-old English monolinguals and the bilinguals ($2 = 0.22$, $df = 1$, $p > 0.5$).

Despite the similarity between the older monolingual group and the bilinguals examined in this study, the overall percentage of null subject use is higher in the latter: while ranges in monolinguals appear to be between 0–11% (average 1.9%), bilinguals resort to them more often (range: 3–17%; average: 8.59%). In addition, the bilinguals and older monolinguals produced null subjects in different pragmatic contexts (Figure 3), suggesting that they are using null subjects for distinct reasons. In contrast with Liceras et al. (2012b), we argue that this study provides initial evidence for a possible role for CLI from Spanish in bilingual English when Spanish is the dominant language. Because of the small quantity of data and participants, our results should of course be interpreted with caution. However, it is conceivable that CLI from Spanish has an impact on what could be considered as a possible transitory stage of null subjects with verbal inflection in English among older four-year-old monolingual children. In this respect, the acquisition of English by bilingual children who are dominant in Spanish could in principle converge with a transitory stage in monolingual English development in which null subjects with inflected verbs are possible.

6. Discussion and implications for future studies

6.1 Analysis

The findings in this exploratory study indicate that while the heritage bilinguals produced a majority of utterances with overt subjects in English (91.4%), they used more null subjects (3–17%, average 8.59%) than age-matched monolinguals, who produced utterances with null subjects 1–6.6% of the time (average: 2.16%). We propose that this pattern is consistent with the Multiple Grammars hypothesis (Amaral & Roeper 2014), namely, that bilinguals have available to them two competing grammars of English: one in which verbal agreement is non-pronominal and cannot license or identify null subjects, and a second one which treats English as a Spanish-like language which licenses null subjects. In other words, these children appear to have set the same values for the Minimal Morphological Threshold in Spanish and a subset of their utterances in English and possibly for the EPP features of the Aboutness Topic. Based on available evidence in the English input, the two grammars are in competition with each other (Yang 2004). We assume that these same grammatical options are available initially to English-speaking children, but that the target-like option eventually outcompetes the Spanish-like grammar, as evidenced by the decline in the number of null subjects produced by English-speaking monolinguals between the ages of 3;0 and 4;0, as well as the decline in the number of pragmatic contexts in which null subjects are produced (Figures 3 and 4). Our proposal is that CLI from Spanish to English causes heritage bilinguals to retain the option to license null subjects for longer than in the case of English monolinguals; thus, it is an indirect form of influence rather than wholesale transfer.

While our results need to be taken with caution due to the limited size of the sample, they indicate command of the syntactic properties licensing null subjects in the Spanish of this group of heritage speakers (ages 4;1–5;3), and possibly CLI from Spanish that affects their patterns of subject acquisition in English. In this regard, our results for heritage learners living in a language contact situation where Spanish is a non-socially dominant language and English is a socially dominant language resemble those reported by Hilles (1991) and Lakshmanan (1994) for L1Spanish/L2 English children more closely than either young L1 English children (around age 3;0) or age-matched Spanish/English simultaneous bilinguals.

On the basis of these results we would like to propose that both the Spanish and a subset of the utterances in the English of the heritage speakers of Spanish analyzed in this study show characteristics of consistently null subject languages. In both languages: (a) null subjects are possible with fully inflected verbs, and (b) they have a similar distribution of pragmatically appropriate antecedents. These results contrast with the findings of studies focused on simultaneous bilingual children arguing

that CLI should occur from the non-null subject language to the null-subject one, but not vice-versa (Hacohen & Schaeffer 2007; Paradis & Navarro 2003; Serratrice et al. 2004; Sorace & Serratrice 2009). In the next section, we identify some key differences between participants in our study and in previous ones.

6.2 Dominance and functionality as key factors influencing CLI

We would like to argue that the (socio)linguistic conditions surrounding the participating children affected the degree and the directionality of the CLI, making their developmental trajectory resemble that of L2 children more closely than the patterns observed in simultaneous bilinguals. Unlike Paradis and Navarro (2003) and Liceras et al.'s (2012a, 2012b), the children in this study constituted a majority of sequential bilinguals (L1 Spanish, L2 English, average age of onset: 3;0). In contrast with Paradis and Navarro's case study, where parents adhered to the one parent/one language strategy, all children in our study received considerably more input in Spanish than in English and were at the early stages of a major shift in their amount of exposure to English, a fact that was reflected in their linguistic complexity in this language and their patterns of language use. Unlike Liceras et al.'s balanced bilingual twins, who did not exhibit crosslinguistic influence in the acquisition of null and overt subjects, our children's dominance in Spanish, in addition to their late consistent exposure to English, seemed to favor CLI effects evidenced as a higher frequency of null subjects with inflected verbs and topic continuity in English.

6.3 Implications for future studies

Despite the limited number of participants in this exploratory study, its findings seem to highlight the importance of considering data from bilingual participants with distinct patterns of input exposure and language use, in this case, from early bilinguals living in a language contact situation, when studying patterns of CLI. In that sense, this initial exploration may contribute to the further understanding of the extent to which CLI in language contact situations, defined as situations in which there are two speech communities involved, is sensitive to the interfaces between different modules and may converge with particular stages of development in monolingual child language. In Spanish, the licensing of null subjects is strongly linked to morphological richness and to topic continuity. Confronted with adult English input that lacks these morphological resources, child bilinguals at the early stages of increased exposure to English seem to hypothesize a morphological mapping for both languages that is compatible with that of adult Spanish grammars, with a preference for null subjects as continuous topics. They appear to exhibit CLI

in morphological mapping from Spanish to English in some of their utterances. CLI in this case may be favored by a possible transitory stage in monolingual English development that allows for null subjects and verbal inflection.

References

Alexiadou, Artemis, and Elena Anagnostopoulou. 1998. Parametrizing AGR: word order, V-movement and EPP-checking. *Natural Language and Linguistic Theory* 16 (3): 491–539. doi: 10.1023/A:1006090432389

Alonso-Ovalle, Luis, Susana Fernández-Solera, Lyn Frazier, and Charles Clifton. 2002. Null vs. overt pronouns and the topic-focus articulation in Spanish. *Italian Journal of Linguistics* 14: 151–170.

Amaral, Luiz, and Tom Roeper. 2014. Multiple grammars and second language representation. *Second Language Research* 30 (1): 3–36. doi: 10.1177/0267658313519017

Argyri, Efrosyni, and Antonella Sorace. 2007. Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism: Language and Cognition* 10 (1): 79–99. doi: 10.1017/S1366728906002835

Austin, Jennifer, Maria Blume, David Parkinson, Reyna Proman, Zelmira Núñez del Prado and Barbara Lust. 1996. Current challenges to the parameter-setting paradigm: The pro-drop parameter. In *Proceedings of the Groningen Assembly on Language Acquisition 1995*, edited by Charlotte Koster and Frank Wijnen. Groningen: Centre for Language and Cognition, 87–97.

Austin, Jennifer, Maria Blume, David Parkinson, Reyna Proman, Barbara Lust, B. and Zelmira Núñez del Prado. 1997. The acquisition of Spanish null and overt pronouns: Pragmatic and syntactic factors. In *Papers on language acquisition: Cornell Working Papers in Linguistics*, edited by Shamita Somashekar, Kyoko Yamakoshi, Maria Blume and Claire Foley 15, 160–177.

Bel, Aurora. 2003. The syntax of subjects in the acquisition of Spanish and Catalan. *Probus* 15 (1): 1–26. doi: 10.1515/prbs.2003.003

Berman, Ruth, and Dan Slobin. 1994. *Relating events in narrative: a crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Blom, Elma, and Sharon Unsworth, eds. 2010. *Experimental methods in language acquisition research*. Vol. 27. Amsterdam: John Benjamins. doi: 10.1075/lllt.27

Bloom, Lois. 1970. *Language development: Form and function in emerging Grammars*. Cambridge: MIT Press.

Bloom, Paul. 1990. Subjectless sentences in child language. *Linguistic Inquiry* 21: 491–504.

Camacho, José. 2013. *Null subjects*. Cambridge University Press. doi: 10.1017/CBO9781139524407

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge: MIT Press.

Chomsky, Noam, and Howard Lasnik. 1977. Filters and control. *Linguistic Inquiry* 8: 425–504.

Contreras, Heles. 1991. On the position of subjects. In *Perspectives on Phrase Structure: Heads and Licensing. Syntax and Semantics* 25, edited by Susan Rothstein. San Diego: Academic Press, 63–79.

Cuza, Alejando. 2012. Cross-linguistic influence at the syntax proper: Interrogative subject-verb inversion in heritage Spanish. *The International Journal of Bilingualism* 17 (1): 71–96. doi: 10.1177/1367006911432619

Frascarelli, Maria. 2007. Subjects, topics and the interpretation of referential pro. *Natural Language and Linguistic Theory* 25 (4): 691–734. doi:10.1007/s11049-007-9025-x

Grinstead, John. 2004. Subjects and interface delay in child Spanish and Catalan. *Language* 80: 40–72. doi:10.1353/lan.2004.0024

Guerriero, Sonia, Amy Cooper, Yuriko Oshima-Takane, and Yoko Kuriyama. 2001. A discourse-pragmatic explanation for argument realization and omission in English and Japanese children's speech. In *Proceedings of the 25th Annual Boston University Conference on Language Development*, vol. 1, edited by Anna Do, Laura Domínguez, and Aimee Johansen. Somerville, MA: Cascadilla Press, 319–330.

Hacohen, Aviya and Jeanette Schaeffer. 2007. Subject realization in early Hebrew/English bilingual acquisition: The role of crosslinguistic influence. *Bilingualism: Language and Cognition* 10 (3): 333–343. doi:10.1017/S1366728907003100

Hilles, Sharon. 1991. Access to Universal Grammar in second language acquisition. In *Point Counterpoint: Universal Grammar in the second language*, edited by Lynne Eubank. Amsterdam and Philadelphia: John Benjamins, 305–338. doi:10.1075/lald.3.14hil

Hinzelin, Marc-Olivier. 2003. The acquisition of subjects in bilingual children. In *(In)vulnerable Domains in Multilingualism*, edited by Natascha Müller. Amsterdam: John Benjamins Publishing, 107–137. doi:10.1075/hsm.1.05hin

Hughes, Mary, and Shanley Allen. 2006. A discourse-pragmatic analysis of subject omission in child English. In *Proceedings of the 30th Annual Boston University Conference on Language Development*, edited by David Bamman, Tatiana Magnitskaia, and Colleen Zaller. Somerville, MA: Cascadilla Press, 293–304.

Hulk, Aafke, and Natascha Müller. 2000. Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition* 3 (3): 227–244. doi:10.1017/S1366728900000353

Hyams, Nina, and Kenneth Wexler. 1993. On the grammatical basis of null subjects in child language. *Linguistic Inquiry* 24 (3): 421–459.

Hyams, Nina. 1986. *Language Acquisition and the Theory of Parameters*. Dordrecht: Reidel. doi:10.1007/978-94-009-4638-5

Hyams, Nina. 1994. Null Subjects in Child Language and the Implications of Cross-Linguistic variation. In *Syntactic theory and language acquisition: Cross-linguistic perspectives*, edited by Barbara Lust, Gabriella Hermon, and Jacklin Kornfilt. Hillsdale, NJ: Lawrence Erlbaum Associates, 289–299.

Hyams, Nina. 2011. Missing subjects in early child language. In *Handbook of language acquisition theory in generative grammar*, edited by Jill De Villiers and Tom Roeper. Amsterdam: Kluwer, 13–52. doi:10.1007/978-94-007-1688-9_2

Juan-Garau, María, and Carmen Pérez-Vidal. 2000. Subject realization in the development of bilingual child. *Bilingualism: Language and Cognition* 3 (3): 173–192. doi:10.1017/S1366728900000328

Lakshmanan, Usha. 1994. *Universal grammar in child L2 acquisition: Morphological uniformity and null subjects*. Amsterdam: John Benjamins Publishing. doi:10.1075/lald.10

Lasnik, Howard. 2001. A note on the EPP. *Linguistic Inquiry* 32 (2): 356–362. doi:10.1162/ling.2001.32.2.356

Liceras, Juana M., Raquel Fernández Fuertes, and Anahí Alba de la Fuente. 2012a. Overt subjects and copula omission in the Spanish and the English grammar of English-Spanish bilinguals: On the locus and directionality of interlinguistic influence. *First Language* 31 (1–2): 88–115. doi:10.1177/0142723711403980

Liceras, Juana M., Raquel Fernández Fuertes, Anahí Alba de la Fuente Genevieve Boudreau, and Elisa Acevedo. 2012b. Interlinguistic influence in simultaneous bilingualism: Core syntax phenomena and lexical transparency. In *Selected proceedings of the Romance Turn IV Workshop*, edited by Sandrine Ferré, Philippe Prévost, Laurice Tuller, and Rasha Zebib. Newcastle upon Tyne: Cambridge Scholars Publishing, 214–239.

López-Ornat, Susana, Sonia Mariscal, Almudena Fernández, and Pilar Gallo. 1994. *La adquisición de la lengua española.* Madrid: Siglo XXI de España.

Marchman, Virginia. 1989. Episodic structure and the linguistic encoding of events in narrative: A study of language acquisition and performance. Unpublished doctoral dissertation, University of California, Berkeley.

Mayer, Mercer. 1969. *Frog, where are you?* New York: Dial Press.

MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk*. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Montrul, Silvina. 2004. Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and Cognition* 7 (2): 125–142. doi: 10.1017/S1366728904001464

Montrul, Silvina. 2009. Incomplete acquisition of tense-Aspect and mood in Spanish heritage speakers. *The International Journal of Bilingualism* 13 (3): 239–269. doi: 10.1177/1367006909339816

Ordoñez, Francisco, and Esthela Treviño. 1999. Left dislocated subjects and the pro-drop Parameter: A case study of Spanish. *Lingua* 107 (1–2): 39–68. doi: 10.1016/S0024-3841(98)00020-5

Orfitelli, Robyn, and Nina Hyams. 2012. Children's grammar of null subjects: Evidence from comprehension. *Linguistic Inquiry* 43 (4): 563–590. doi: 10.1162/ling_a_00106

Otheguy, Ricardo, Ana Celia Zentella, and David Livert. 2007. Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language* 770–802.

Paradis, Johanne, and Samuel Navarro. 2003. Subject realization and crosslinguistic interference in the bilingual acquisition of Spanish and English: what is the role of the input? *Journal of Child Language* 30 (2): 371–393. doi: 10.1017/S0305000903005609

Park, Heejung, Kim Tsai, Lisa Liu, and Anna Lau. 2012. Transactional associations between supportive family climate and young children's heritage language proficiency in immigrant families. *International Journal of Behavioral Development* 36 (3): 226–236. doi: 10.1177/0165025412439842

Pérez-Leroux, Ana Teresa, Alejandro Cuza, and Danielle Thomas. 2011. Clitic placement in Spanish–English bilingual children. *Bilingualism: Language and Cognition* 14 (2): 221–232. doi: 10.1017/S1366728910000234

Picallo, M. Carme. 1998. On the Extended Projection Principle and null expletive subjects. *Probus* 10 (2): 219–242. doi: 10.1515/prbs.1998.10.2.219

Pierce, Amy. 1992. *Language acquisition and syntactic theory: A comparative analysis of French and English child grammars*. Dordrecht: Kluwer. doi: 10.1007/978-94-011-2574-1

Pires, Acrisio, and Jason Rothman. 2009. Disentangling sources of incomplete acquisition: An explanation for competence divergence across heritage grammars. *International Journal of Bilingualism* 13 (2): 211–238. doi: 10.1177/1367006909339806

Renner, Judith. 1988. Development of temporality in children's narratives. Unpublished doctoral dissertation, University of California, Berkeley.

Rizzi, Luigi. 1993. Some notes on linguistic theory and language development: The case of root infinitives. *Language Acquisition* 3 (4): 341–393. doi: 10.1207/s15327817la0304_2

Rizzi, Luigi. 2005. On the grammatical basis of language development: A case study. In *The Oxford handbook of comparative syntax*, edited by Guglielmo Cinque and Richard Kayne. New York: Oxford University Press, 70–109.

Sano Tetsuya, and Nina Hyams. 2000. Agreement, finiteness, and the development of null arguments. In *Acquisition of scrambling and cliticization*, edited by Susan Powers and Cornelia Hamann. Dordrecht: Kluwer, 345–396.

Sebastián, Eugenia. 1989. "Tiempo y aspecto verbal en el lenguaje infantil." Unpublished doctoral dissertation, Madrid: Universidad Autónoma de Madrid.

Serratrice, Ludovica, Antonella Sorace, and Sandra Paoli. 2004. Crosslinguistic influence at the syntax-pragmatics interface: subjects and objects in Italian-English bilingual and monolingual acquisition. *Bilingualism Language and Cognition* 7 (3): 183–205. doi:10.1017/S1366728904001610

Serratrice, Ludovica. 2006. Referential cohesion in the narratives of bilingual English-Italian children and monolingual peers. *Journal of Pragmatics* 39 (6): 1058–1087. doi:10.1016/j.pragma.2006.10.001

Shin, Naomi, and Ricardo Otheguy. 2009. Shifting sensitivity to continuity of reference: subject pronoun use in New York City. In *Español en los Estados Unidos y otros contextos de contacto: Sociolingüística, ideología y pedagogía*, edited by Manuel Lacorte and Jennifer Leeman. Madrid/Frankfurt: Iberoamericana/Vervuert, 111–136.

Slabakova, Roumyana, and Ivan Ivanov. 2011. A more careful look at the syntax–discourse interface. *Lingua* 121 (4): 637–651. doi:10.1016/j.lingua.2010.05.003

Sorace, Antonella, and Ludovica Serratrice. 2009. Internal and external interfaces in bilingual language development: beyond structural overlap. *International Journal of Bilingualism* 13 (2): 195–210. doi:10.1177/1367006909339810

Sorace, Antonella, Ludovica Serratrice, Francesca Filiaci, and Michela Baldo. 2009. Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua* 119 (3): 460–477. doi:10.1016/j.lingua.2008.09.008

Suñer, Margarita. 2003. The lexical preverbal subject in a Romance Null Subject Language. In *A Romance perspective on language knowledge and use: Selected papers from the 31st Linguistic Symposium on Romance Languages (LRSL), Chicago, 19–22 April 2001*, vol. 238, edited by Rafael Núñez-Cedeño, Luis López, and Richard Cameron. Amsterdam and Philadelphia: John Benjamins Publishing, 341–358. doi:10.1075/cilt.238.25sun

Valian, Virginia. 1990. Null subjects: A problem for parameter-setting models of language acquisition. *Cognition* 35 (2): 105–122. doi:10.1016/0010-0277(90)90011-8

Valian, Virginia. 1991. Syntactic subjects in the early speech of American and Italian children. *Cognition*, 40 (1–2): 21–81. doi:10.1016/0010-0277(91)90046-7

Villa-Garcia, Julio. 2012. Spanish subjects can be subjects: Acquisitional and empirical evidence. *Iberia: an International Journal of Theoretical Linguistics* 4 (2): 124–169.

Yang, Charles. 2004. Universal Grammar, statistics or both? *Trends in cognitive sciences* 8 (10): 451–456. doi:10.1016/j.tics.2004.08.006

Yip, Virginia, and Stephen Matthews. 2000. Syntactic transfer in a Cantonese-English bilingual child. *Bilingualism: Language and Cognition* 3 (3): 193–208. doi:10.1017/S136672890000033X

Zubizarreta, María Luisa. 1999. Word order in Spanish and the nature of nominative case. In *Beyond principles and parameters:Essays in memory of Osvaldo Jaeggli*, edited by Kyle Johnson, and Ian Roberts. Dordrecht: Kluwer Academic Publishers, 223–250. doi:10.1007/978-94-011-4822-1_9

CHAPTER 11

Return to Frenchville

Tracing a near-merger from legacy data

Barbara E. Bullock and Jenna Nichols
The University of Texas at Austin

This study investigates a potential case of near-merger in legacy French data from Frenchville, Pennsylvania. With previous research having found a robust differentiation in production between French schwa and the front mid rounded vowels by the final generation of Frenchville French speakers, we hypothesize that a former near-merger of these vowels enabled a subsequent demerger. The present analysis examines schwa and the mid vowels in interview data from the penultimate generation of Frenchville speakers and finds evidence for a near-merger, as the vowels are similar but not identical in duration and spectral quality. The data also support the notion that the differentiation of these vowels by the final generation was likely an innovation, rather than the result of transmission.

Keywords: near-merger, language change, legacy data, rhoticization, vowel systems, schwa, language isolate

1. Introduction

In an analysis of the phonological characteristics of the French spoken by the last two speakers of French from the village of Frenchville, Pennsylvania (USA), Bullock and Gerfen (2005) demonstrate that the front mid rounded vowels of French /ø, œ/ have converged to an American English-like rhoticized schwa [ɚ], which means that French words like *deux* 'two' and *neuf* 'nine' have a rhotic perceptual quality and are heard as [dɚ] and [nɚf], very close to the pronunciation of words such as *her* [hɚ] and *surf* [sɚf].[1] At the same time, the authors show that their participants' productions of schwa in French resist rhoticization so that the vowel of the preposition *de*, when not elided, retains the percept of a front, rounded vowel, and thus

1. The English of Frenchville, PA, located in the US Midland dialect zone, is a rhotic variety.

DOI 10.1075/rllt.11.11bul

contrasts audibly with the vowel in *deux* 'two' ([dø] *de* vs. [də̞] *deux*). A similar situation occurs in Gatineau, Quebec where /ø, œ/ can be produced as [ə̞] among some French speakers (Mielke 2013) but schwa cannot (Mielke, p.c.). The fact that schwa is phonetically distinct from /ø, œ/ in these communities is remarkable given that much of the phonological literature devoted to the status of schwa in French holds that schwa has an overlapping phonetic distribution with /ø, œ/ (Morin 1978; Tranel 1987; Walker 1993). That is, schwa appears to have phonetically merged with these vowels. But if schwa were indeed phonetically non-distinct from [ø] or [œ] then, as pointed out by Bullock and Gerfen, it remains to be understood why schwa resists undergoing the rhoticization that has impacted /ø, œ/ in Frenchville (and, likewise, in Gatineau).

Because this rhoticization had not been observed among linguists in earlier investigations of Frenchville French (FF) speakers (Valdman 1979; Mougeon and Uritescu 2006), Bullock and Gerfen (2005) suggest it is an innovation of the last generation of speakers, potentially brought about by the speakers establishing a perceptual match between English [ə̞] and [ø, œ], all phonologically unreduced vowels that are produced with lip rounding. They posit that a plausible explanation for the apparent demerger of the putatively identical phones in FF is that schwa and the mid, front rounded vowels were not completely neutralized in the input variety acquired by the speakers but were potentially phonetically distinct at a fine-grained sub-phonemic level. If the vowels were not fully overlapping in their categories but were perceived to be, this would constitute a case of *near-merger* such as those documented by a long line of sociolinguists with regard to varieties of English (Labov, Yeager & Steiner 1972; DiPaola & Faber 1990; Labov, Karen & Miller 1991; Labov 1994; Faber & DiPaolo 1995). Under this view, the last generation of FF speakers would be said to have *de-merged* or reversed the near-merger.

In order test the hypothesis of a potential reversal of a near-merger in Frenchville, the present work analyzes legacy data from two speakers who were recorded in 1988 and whose speech is taken to represent the variety acquired by the speakers investigated by Bullock and Gerfen (2004a, 2004b, 2005). In particular, we investigate here the possibility that the rhoticized front mid vowels of FF emerged from an input variety that was non-rhotic but in which schwa was phonetically distinct from /ø, œ/ along at least some acoustic dimensions, understood as spectral and durational. In particular, we examine the productions of two FF speakers who are approximately 30 years older than those studied by Bullock and Gerfen, and who can be taken to represent the previous generation of French-English bilingual speakers in that community. We should note that these speakers do not produce saliently rhotic vowels in their French. From the speech of this older generation, we attempt to confirm whether their productions of [ø, œ] (taken here as a single category for reasons to be explained below) and schwa show a complete merger, an

incomplete merger, or whether they already show two distinct categories. Before we present the acoustic analyses, we overview the status of schwa in French and the history of the language isolate community of Frenchville. We then describe the legacy data upon which the analysis is based and trace the relation among the speakers analyzed here and those analyzed in Bullock and Gerfen (2005).

2. Background: The status of schwa French

The categorization of schwa in the sound system of French (understood as *le français de référence* (Morin 2000)) remains a matter of controversy in large part because of its apparent overlapping phonetic distribution with the vowels /ø, œ/. There are linguists who hold that, when realized, schwa has a phonetic quality that vacillates between phonetic [ø] and [œ] (Delattre 1966) and other phonologists who associate its pronunciation to the timbre of either one or the other of these vowels (Morin 1978; Walter 1982; Verluyten 1988). Given the close auditory and acoustic similarity of schwa to the other front rounded vowels, debate in the phonological literature has focused mainly on whether schwa, traditionally represented as /ə/, should be considered as categorically separate from /ø, œ/ (Léon 2007; Coveney 2001), whether it should be considered an elidable allophone of the mid front rounded vowels (Walker 1993) or whether it is a non-phonemic "lubricant" vowel that serves to break up clusters of consonants (Martinet 1972).

The synchronic basis for categorizing schwa as distinct from the mid front rounded vowels generally rests on its orthographic representation and on its distributional properties rather than on its acoustic identity. With regard to orthography, the latter vowels, which derived from historical diphthongs, are represented by the digraph <eu> while schwa is almost invariably represented by <e>, referred to as *e-muet* (silent), *e-instable* (unstable), or *e-caduc* (fallen) after its propensity to delete.[2] With respect to its distributional properties, schwa patterns differently from other vowels in several major ways. First, it is the only vowel that can be deleted whenever its elision would not result in an overly complex cluster of consonants (see Léon 2007 for an overview). Conversely, schwa is the only vowel that can be inserted with no change of semantic content as an excrescent vowel word internally and finally (e.g. *capter* 'to get', pronounced [kapəte] rather than [kapte]; *film*, pronounced [filmə] (see Bazylko 1981; Colantoni & Steele 2011; Hansen 1994, 2012). Unlike other vowels, schwa is restricted from occurring in word-initial position and it also cannot occur in a stressed syllable unless it falls in enclitic position: *prends-le!*

2. There is but a handful of words where an elidable schwa is represented in a different fashion: *faisons, monsieur,* and *peut-être*.

'take it!' Additionally, schwa can occur in hiatus when it precedes another vowel phonetically, generally in the context of *h-aspiré,* the abstract segment that blocks elision, as in *le hibou* (cf., **l'hibou*) [ləibu] 'the owl', but it cannot follow another vowel as other vocalic segments do (e.g. *prieur* [pʁiœʁ] 'religious prior'). Finally, in stem modifying verb paradigms, unstressed schwa enters into a morpho-phonological alternation with stem stressed [ɛ] and not with stressed [ø,œ] (e.g. *mener* 'to lead' [məne] ~ *mène* [mɛn]).

While phonologists have sought the proper characterization of schwa based on these divergent distributional properties, phoneticians have examined the spectral and durational manifestations of schwa. In an early work on natural speech, Malécot and Chollet (1977) discovered schwa to be more diffused in phonetic space than the other mid front rounded vowels with a quality that most resembles that of the closed mid vowel [ø]. Bazylko (1981) conducted an instrumental analysis of the front mid round vowels, including schwa, and remarked that they were acoustically distinct even though they were not distinguishable to the listener, a hallmark of near-mergers. More recently, Fougeron, Gendrot, and Bürki (2007) examined the phonetic realizations of /ø/, /œ/, and /ə/ from 574 speakers in a French broadcast news corpus, finding durational and spectral differences between schwa and the two front vowels despite considerable overlap in the distribution of these vowels. Overall, they determine that schwa is characterized as shorter than other vowels, intermediate in height (F1), with a higher F2 than [ø, œ], and an intermediate degree of rounding (F3). Their study confirms the observations of Malécot and Chollet (1977) that schwa is closer to [ø] than to [œ].

In a study of the same French broadcast corpus, Bürki et al. (2011) investigate phonetic reduction in schwa. From the vowels labeled schwa, the authors find a range of durations from 8 ms to 150 ms (with a mean of 50 ms); shorter schwas were more likely to have a higher F2, that is, to be more fronted than longer ones. In a follow-up perception experiment, the authors find that listeners disagree about the presence or absence of schwa when the vowel is less than 50 ms, indicating that the presence of a schwa of relatively short duration can be confused with its absence. In a perceptual confusability study of French vowels, Hall and Hume (2013) find the highest rates of identification errors among the mid front rounded vowels [ø, œ, ə] and confirm the tendency noted in Bürki et al. (2011) for the absence of a vowel to be confused mostly with schwa. Hall and Hume find that schwa is the most accurately identified of the three mid front vowels but also that it is identified as [ø] as often as it is identified correctly. By contrast, tokens of [ø] and [œ] are more often confused with one another than they are correctly identified.

The production studies reviewed above indicate that most speakers of contemporary hexagonal French produce a schwa that is acoustically distinct from the other front round mid vowels in duration and spectra despite some overlap

between the categories. In fact, nine out of the subset of twelve speakers analyzed individually by Fougeron et al. (2007) distinguished schwa from both /ø/ and /œ/. At the same time, it appears to also be the case that there is a high rate of perceptual confusability among these vowels among native speakers. This suggests a scenario of near-merger in which speakers produce but cannot discriminate fine-grained phonetic distinctions between words (Labov, Yeager & Steiner 1972; Janson & Schulman 1983; DiPaola & Faber 1990; Labov, Karen & Miller, 1991; Labov 1994; Faber & DiPaolo 1995; Yu 2007).

2.1 A case for near-merger?

Outside the sociolinguistic literature, accounts of near-merger in linguistics are controversial because they imply that speakers are capable of maintaining fine-grained differences in their speech production that they, and the linguists who observe their behavior, are unable to perceive (Manaster Ramer 1996). For instance, Hickey (2004) declares that a near-merger could only be transitory and that it could never be transmitted to new generations because children would have to acquire phonetic distinctions that they would not be able to hear. Under his view, the potential for a near-merger to be reversed into a new or a renewed contrast would be impossible in the absence of external pressures such as orthography or contact with a dialect that maintains a robust phonemic distinction.

However, the phonetic and perceptual evidence reviewed above suggests that the well-studied case of schwa and its mid front round counterparts in French could be considered to constitute a case of near-merger. Most speakers of contemporary French maintain systematic phonetic distinctions between schwa and /ø, œ/ that they are unable to consistently categorize in perception. If the phonetic system of the French imported to Frenchville in the 19th century was at all similar to the contemporary French studied by Fougeron et al. (2007), then the behavior of the speakers of FF studied by Bullock and Gerfen (2005) might provide evidence that the reversal of a near-merger in the absence of pressure from orthography or from a related dialect where the contrast is maintained may indeed be possible because the phonetic reflex of /ø, œ/ in the speech of the FF speakers studied has diverged *dramatically* from that of schwa.

3. Participants and data

The French speakers who populated Frenchville arrived from the Haute-Marne and Haute-Saône departments of France in two waves of immigration in the early to

mid 19th century to settle along the banks of the West Branch of the Susquehanna River as loggers and farmers. Frenchville was geographically isolated and to this day remains sparsely inhabited. Its Catholic inhabitants of the 19th and early 20th century, finding themselves surrounded by Protestants, tended to remain endogamous. This, and the fact that they had a one-room schoolhouse in which they educated their children locally allowed them to maintain their French-English bilingualism well into the 20th century. This means that the residents of Frenchville surpassed the normal pattern of shift by which immigrant communities become monolingual in English by the third generation. The individuals whose recordings constitute the data for this study are third generation francophones on the side of their maternal grandmother but likely fourth or fifth generation French descendants on their paternal side.

The data for the present study are provided by a recorded conversation between a brother and a sister who were interviewed by a linguist, with his wife as a participant, in 1988. Thus, the data for the present study constitute *legacy data* in the sense that the recording was made in the past rather than recorded in the field for the purposes of a new study (Bounds et al. 2011). The data, which was originally in analog format on cassette tape, were given to the first author by the linguist and are used with the permission of the heir of the individuals interviewed. Technicians at the University of Texas Sound Lab converted the recording to a digital format with a sampling rate of 44.1 kHz, 16-bit quantization. Given the nature of the recordings – a roughly 120 minute conversation, partly in French and partly in English, between four people in a relatively noisy environment – the observations upon which we can report are limited. Nonetheless, we are fortunate to have this legacy data as it affords us the possibility of investigating the potential for variation and change over time.

The speakers analyzed for the present study are siblings, a male, FP (b. 1909) and female, EV (b. 1903). Importantly, the siblings analyzed in this study are first cousins of the brothers studied by Bullock and Gerfen, who were born in 1930 and 1933, a generation later than the individuals analyzed here. As indicated by Bullock's field notes, FP and EV were well known to the brothers and, according to one of the brothers (NE), they always spoke French with their francophone relatives and neighbors in Frenchville, as indicated by the following exchange, which specifically questioned their interactions with the family of FP and EV:

Bullock	*Donc vous parliez le français avec eux.*
NE	*Ah oui.*
Bullock	*Mais il y avait d'autres personnes qui parlaient anglais…?*
NE	*ah…nnnn…on parlait le français. Si on parlait…s'ils pourraient* [sic] *parler le français, on parlait le français*!

Bullock So you spoke French with them?
NE Ah yes.
Bullock But if there were other people who spoke English…?
NE Ah…nnnnaw…we spoke French. If they spoke…if they could speak French, we spoke French!

Like their younger cousins, FP and EV learned French at home, not in school. Thus, they, too, were not literate in French although EV claimed to have taught herself to read from a book owned by her French-born grandmother. Of the four Frenchville speakers, then, the orthographically distinct representations for schwa and /ø, œ/ were potentially only available to EV.

Both FP and EV had traveled to France. FP served as an interpreter for a short time in Marseilles during WWII and EV once traveled to France as a tourist (Bullock, field notes). Thus, they each experienced a single, relatively short contact experience with French speakers outside the context of Frenchville. In this way, they are distinguished from their cousins, NE and KE, who had very little exposure to French speakers outside the occasional visit from a curious linguist or a tourist.

4. Methods

The methods of coding for the present analysis closely follow those of Bullock and Gerfen (2005) in collapsing the mid front vowels /ø, œ/ into a single, merged category. This decision is based on the fact that these vowels are only potentially contrastive in a few minimal pairs (e.g. *jeune* [œ] 'young' versus *jeûne* [ø] 'a fast') in any variety of French; otherwise, they are found in complementary distribution with [ø] appearing in open syllables and [œ] in closed syllables (except for those closed with non-nasal coronal consonants in which case the vowel is said to raise (e.g. *creuse* 'to dig' [krøz] (Andreasson & Østby 2014))). In FF, the reflex of /ø, œ/ for the later generation of speakers was shown to be rhoticized irrespective of the structure of the syllable in which it is found and it is the rhotic/non-rhotic distinction that is of interest to this analysis. The distributional facts of /ø, œ/, then, militate in favor of considering these vowels to be a single category in FF particularly as there would only ever be a small set of words in any French corpus for which separate lexical categories of /ø/ versus /œ/ would appear. In a corpus of the size we consider here, these simply do not appear. For the purposes of coding the vowel productions, we considered any word spelled with <eu> as falling into the merged mid front round vowel category, which we label "œ" for convenience. By contrast, any vowel that is spelled with <e> and occurs between consonants (e.g. *petit*) or within one of the

nine monosyllabic functional elements (*je, te, se, ce, me, le, que, ne, de*) is considered to be in the schwa category.

For the acoustic analysis, all tokens of schwa and of the mid front rounded vowels were extracted from the interview via Praat (Boersma & Weenik 2013). The beginning and endpoint of each vowel was subsequently delineated on a textgrid with the interval corresponding to the vowel portion only. The intervals were labeled with the word containing the vowel. The onset/offset of the vowels were determined by an examination of F2 and of abrupt changes in periodicity and intensity in the waveform. All onsets and offsets were marked at zero crossings. If one of the highlighted vowels was unintelligible or had no discernible formant structure, it was not included in the measurements. Formant measures were extracted automatically using an original Praat script written by the second author and spot-checked manually for accuracy and for outliers.

Following the procedures of Bullock and Gerfen (2005), who found a correlation between the strong percept of the North American [ɚ] and a steep drop in F3 among their participants and of Mielke (2013), we interpret a dramatic drop in F3 as a potential acoustic indicator of vowel rhoticization. We used a vowel script to take formant measures of F1, F2, and F3 at the midpoint and at the 3/4 point of the vowel interval. These measures were converted to Bark, a psycho-acoustic scale, in order to more adequately represent audible differences. These two particular points were selected in order to avoid changes in F2 due to formant transitions caused by co-articulation of the vowel with neighboring consonants. The decision to measure two points follows from the need to examine the potential rhoticity of the vowels in terms of changes to F3. The spectral reflex of rhoticity is reflected in a falling trajectory and low offset locus for F3. In other words, rhoticized vowels should show a steep drop in F3 from the midpoint to the 3/4 point. Therefore, rhoticity can be quantified by the simple formula: $F3_{(3/4point)} - F3_{(midpoint),}$ where larger negative values (in terms of absolute value) equal larger drops in F3 and greater rhoticity. We hypothesize that, unlike for the younger subjects surveyed in Bullock and Gerfen (2004b, 2005), there is no significant degree of rhoticity for the [œ] vowels of FF among the older speakers surveyed in the legacy data.

5. Results

Given that we are analyzing legacy data that was not recorded in ideal conditions, our observations are somewhat limited in number. They are also skewed in that the older, female participant, EV, spoke more than her younger brother, FP. This was also the case with their cousins analyzed by Bullock and Gerfen (2005); the older sibling spoke much more than the younger one. Nevertheless, we were able to

isolate an equivalent number of observations to match those of Bullock and Gerfen (2005). The descriptive token counts are shown in Table 1.

Table 1. Total number of tokens

Speaker	[œ]	schwa
EV	80	116
FP	28	36

There are more tokens of schwa than [œ] for each sibling, which reflects the distribution of the vowels in the lexicon of French (Malécot 1974). While these token counts are small, particularly for FP, they are nearly exhaustive, representing the extent of the data that we have available. However, given the small count numbers for FP, the results of the statistical tests performed on his productions that follow can only be interpreted tentatively.

5.1 Results of F3 drop

To test the hypothesis regarding the (non)-rhoticity of the [œ] vowels in FF, we ran two-tailed, non-paired t-tests for each speaker separately. These compare the mean change in F3 from the midpoint to the 3/4 point for schwa and for [œ], where rhoticization is implicated with greater degrees of F3 drop. Negative numbers (i.e., a lower value for F3 at the 3/4 point than at the midpoint) would indicate rhoticity. Note that, for FP, there is no significant difference in the value calculated for mean drop in F3. Further, and more important, the values are positive, indicating that F3 *rises* slightly from the mid-point to the 3/4 point of both vowels in FP's speech. The data for his sibling, EV, is provided in Table 3.

The pattern here is identical to that of speaker FP. Again, the values are positive, so F3 rises slightly rather than falls over the course of the vowel from the midpoint to the 3/4 point with no statistically significant differences between the two vowels. Thus, the statistical tests for rhoticity reveal that there is no significant difference in the putative drop in F3 between schwa and lexical [œ], a fact which is confirmed auditorily. Thus, these speakers do not produce vowels that acoustically reflect potential rhoticity nor do they sound rhoticized either for schwa or lexical [œ].

Table 2. Mean change in F3 for FP (*non-significant*)

Vowel	Mean change in F3
[œ] (*n*=28)	0.13 Bark
schwa (*n*=36)	0.17 Bark

Table 3. Mean change in F3 for EV (*non-significant*)

Vowel	Mean change in F3
[œ] (*n*=80)	0.13 Bark
schwa (*n*=116)	0.06 Bark

An additional result relevant to the measures of the third formant is that there were no significant differences between F3 in [œ] and schwa at any measurement point. As F3 correlates with lip rounding, as well as with rhoticization, this means that both vowels are rounded to the same degree for both speakers.

5.2 Results for spectra of the vowels: F1

The results from the F3 measures indicate that there are no rhoticized vowels present in the speech of FP or EV. To examine the acoustic similarities or differences between the two categories of vowels, we measured the values for the first three formants for each speaker separately at each measurement point and compared these measures between the vowels coded as schwa and those coded as [œ]. Since we have no prediction regarding whether these vowel categories differ significantly in their spectral properties, we employed two-tailed tests within each formant and measuring point. Comparisons were made for each speaker of F1 through F3 between the two vowels at the midpoint and at the three-quarter point (i.e. six tests for each speaker). For speaker FP, the two vowel categories were significantly different only in F1 at the midpoint of the vowel. This is shown in Table 4. Notice that schwa has lower mean F1 value than [œ] for speaker FP. This means that the schwa category is slightly higher in the vowel space than the category for [œ].

Table 4. F1 at midpoint for FP ($t = -1.676$, $df = 64$, $p = .049$)

Vowel	Mean F1
[œ] (*n*=28)	4.26 Bark
schwa (*n*=36)	3.98 Bark

The analysis for EV likewise reveals that, of all the comparison points, the only significant difference between schwa and [œ] for EV was found at the midpoint of F1, as shown in Table 5.

Table 5. F1 at midpoint for EV ($t = -1.717$, $df = 194$, $p = 0.044$)

Vowel	Mean F1
[œ] (n=80)	4.44 Bark
schwa (n=116)	4.27 Bark

Notice that, like her brother, EV's schwa has, on average, a lower F1 than [œ], indicating that schwa is higher. This significant height distinction among schwa and [œ] at the midpoint in both speakers appears to level out at the 3/4 point of the vowel.

5.3 Results for the spectra of vowels in the front-back dimension: F2-F1

In order to examine the acoustic dimension that correlates with the perceived backness of a vowel, we calculated the speakers' F1 and F2 for both [œ] and schwa at midpoint. Larger F2 values indicate relative frontness while smaller values indicate more posterior vowels. The two-tailed, unpaired t-tests indicated no statistical difference in the backness dimension for schwa and [œ] for either speaker at the 95% confidence interval. The vowels for EV, for whom we have more observations and who appears to show differences in height and backness between these vowels, are plotted in traditional vowel space in Figure 1 with the y-axis representing vowel height (inversely correlated with F1) and with the value for F2, plotted on the x-axis to represent vowel backness. Schwa vowels are represented as open circles and [œ] as filled diamonds. We see that there is overlap among the two categories

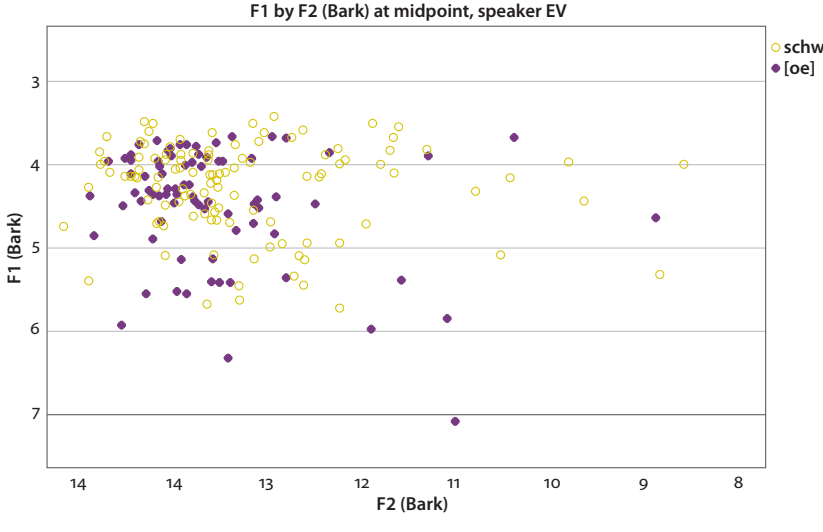


Figure 1. Vowel scatterplots for speaker EV

and some lower, centralized outliers for both vowels. However, there are patterns to be discerned as well. For EV, schwa appears to be more diffuse in the vowel space and a bit higher overall while [œ] is more compact and clustered more toward the mid-range of height in the anterior portion of the vowel space.

5.4 Results of durational measures

In a final analysis, we examine the durational differences between schwa and lexical [œ] for each speaker. To normalize vowel length for each speaker, we followed the grand mean procedure developed by Wassink (2006), where the "grand mean" is an average of the speaker's average duration for each phoneme produced. The normalized duration of a token is calculated by subtracting the grand mean from the duration of the individual token. For each speaker, we compared these normalized durations between the productions of [œ] and schwa. The results of the two-tailed t-tests are shown in the tables below.

Table 6. Duration results for FP ($t = -5.128$, $df = 64$, $p < 0.001$)

Vowel	Average duration
[œ]	194 ms
schwa	92 ms

The data for the male speaker, FP, demonstrates the he maintains a length distinction between schwa and lexical [œ] that is highly significant. On average, his [œ] vowels are twice as long than the vowel he produces for schwa. Crucially, this durational difference cannot be due to positional factors. Stressed, phrase-final schwa can only be found under rare conditions and all the schwa tokens measured here are in phrase-internal position. Although one might expect [œ], a vowel that occurs in tonic position in French, to fall phrase-finally more frequently than schwa and, consequently, be subject to the lengthening that is typical in that position in French, all but two of the tokens of [œ] in FP are phrase-internal. These two tokens do not disproportionately affect the outcome, for even by excluding them, a t-test still reveals a significant durational difference at the $p < 0.01$ level.

Table 7. Duration results for EV ($t = -1.695$, $df = 194$, $p = 0.046$)

Vowel	Average duration
[œ]	119 ms
schwa	99 ms

By contrast, the durational results for EV in Table 7 show only slightly significant length differences between [œ] and schwa. Her schwa is the shorter of the two on average but she does not approximate the magnitude of the length differences in these categories that is shown by her brother, who appears to maintain a clear length distinction between his tokens of schwa and [œ].

6. Discussion

The results of this study produce several results bearing on the questions of whether schwa and [œ] are distinct categories or merged ones in the speech of this generation of FF speakers. First, the results demonstrate incontrovertibly that there is no rhoticization present in this generation of speakers. This means that there is no evidence that the vowels of these speakers pattern in separate rhotic vs. non-rhotic distributions as they did for the later generation of speakers studied by Bullock and Gerfen (2005) in which the mid front rounded vowels, but not schwa, were strongly rhoticized. This finding supports the notion that rhoticization was likely an innovation in the speech of the younger generation and not the result of transmission (in the sense of Labov 2007).

While we can state that the vowels under analysis here do not constitute two separate categories with regard to rhoticity, the results of this study demonstrate that both speakers make fine-grained phonetic distinctions between [œ] and schwa and they tend to make such distinctions in the same direction although not to the same degree. Both speakers produce [œ] with significantly longer duration on average than schwa and, on average, both speakers produce [œ] with a significantly higher F1 than schwa, locating it slightly lower in vowel space. The durational results match the phonetic studies of contemporary spoken French in France, where schwa is found to be shorter on average than /ø, œ/ (Fougeron et al. 2007; Bürki et al. 2011; Malécot and Chollet 1977). In the present case, FP, the male speaker appears to make very large differences in duration between these vowel categories, although, as mentioned above, the low number of observations for this speaker entails that the statistical results for his productions must be treated with caution. However, his older sister, EV, also maintains length distinctions between the vowels in the same direction, producing longer [œ] vowels than schwa.

The spectral results of the present study indicate that both speakers maintain distinct categories for schwa and [œ] with regard to vowel height. Because of our low token count, and in the interest of comparing our results to those of Bullock and Gerfen (2005), who collapse the mid front round vowel category, we did not distinguish [œ] and [ø], as French phoneticians have done. Therefore, we cannot compare our results directly to the values for contemporary spoken French

as measured separately for [ə, ø, œ]. Nonetheless, our result for vowel height, in which schwa is higher than the [œ] vowels, is consistent with those of Malécot and Chollet (1977) and of Fougeron et al. (2007) where schwa was shown to map more closely to the higher variant, [ø], than to its open counterpart when these vowels are considered separately. Interestingly, the inter-speaker analysis of twelve individual speakers selected from the broadcast corpus by Fougeron et al. (2007) indicated that 6 of them differentiated optional schwa from the open vowel /œ/ by F1 alone and the other 6 by F1 in combination with either F2 (backness) or F3 (rounding). This, coupled with our results, might indicate that F1 offers a fairly consistent cue in French for distinguishing optional schwa from the open variant /œ/. In this respect, it bears mentioning that our results are convergent with those found for the later generation of FF speakers where schwa was found to be produced consistently higher than their rhoticized reflex of lexical [œ].

The spectral analysis of backness (F2-F1) revealed that there were no statistically significant differences between schwa and [œ] for backness, although there was a trend for the female speaker to produce schwa as more centralized than [œ]. Similarly, Bullock and Gerfen (2005) found that their speakers produced largely overlapping categories for these vowels in the front – back dimension although there were trends toward backing lexical [œ], rather than schwa, among the later generation. Of course, it must be remembered that these speakers were producing rhoticized variants, [ɚ], for the [œ] class of words. Rhotic vowels are produced with lip-rounding, which results in a drop in F2 (and F3). Such a gesture backs the vowel and decreases the F2-F1 difference. Since the earlier generation of FF under analysis here produces non-rhotic variants for the [œ] class of words, no such backing would be expected.

The results of the legacy data analyzed in this study are consistent with the claims of Bullock and Gerfen (2005) that the brothers could have acquired a variety of French in which schwa and [œ] were not completely merged at the sub-phonemic level. The present results show that their older relatives show consistent differences between schwa and [œ] in both duration and vowel height that are similar to those that have been found among contemporary French speakers where schwa remains distinct from the other mid front rounded vowels despite considerable inter-speaker variation.

When we consider the social embedding of these vowels in Frenchville, it is remarkable that the language shares these fine-grained phonetic distinctions with other varieties of contemporary spoken French. The French speakers of Frenchville were isolated from communities of monolingual French speakers for at least two generations before these legacy recordings were made. In addition, the residents of this community were largely only literate in English, not French, so the orthographic differences that potentially help to reinforce phonetic distinctions among these

vowels in other communities were unavailable to the speakers of Frenchville surveyed in this study and in previous studies. Finally, speakers in Frenchville were always under social pressure to use English in order to participate in the socio-economic, political, educational, and religious institutions of central Pennsylvania, and their French gradually yielded to English dominance with successive generations. Nonetheless, the same sub-phonemic properties that distinguish schwa from the mid front round vowels in France today appeared to have also been transmitted down to the speakers analyzed here.

Bullock and Gerfen (2005) had speculated that if such sub-phonemic differences had existed in the input to the brothers that they studied, they could have remapped the [œ] class to [ə] potentially due to establishing a perceptual equivalence for it to the English vowel but maintained the schwa as a distinct mid front (non-rhoticized) round vowel. In other words, they demerged them potentially under the effect of contact from English, not from contact with a non-merged French variety. The legacy data that we have analyzed here provides support for the notion that schwa and [œ] in the input to these speakers constituted a case of near merger and not neutralization. Although the acoustic properties of the vowel classes of the speakers analyzed are very similar and largely overlapping, they do constitute separate categories in production. How and why the youngest generation remapped these vowels into separate rhotic and non-rhotic categories remains a mystery, although perhaps the speakers' bilingualism influenced this split. At the very least, the results of this study suggest that the reversal of a near-merger is not only possible, it is attested.

7. Conclusion

We have argued that the data analyzed here reflects a situation of phonetic near-merger of schwa and [œ] in the French spoken by Frenchville speakers born at the beginning of the 20th century. That is, the vowels were acoustically similar in duration and spectral quality but not fully neutralized. This near-merger may have facilitated the demerger of the categories by the subsequent generation of speakers analyzed by Bullock and Gerfen (2004a, 2004b, 2005). Thus, we take the position, contra Hickey (2004) and Manaster Ramer (1996), that the reversal of near-mergers in the absence of the external pressure of orthography is possible. We cannot rule out the possibility that external contact might play a role in reversing near-merger. However, in the Frenchville French case, this would not entail contact with another French variety that maintains the contrast but contact with another language (English) in which a similar schwa-like vowel and a similar mid rounded vowel (albeit, rhoticized) constitute separate categories. The role of cross-linguistic contact

in the reversal of an apparent merger has not, to our knowledge, been previously attested. We cannot make a strong case for it here because our data are sparse and because the appearance of rhotic vowels in Gatineau French appear to result from a change from below rather than from convergence with English (Mielke 2013). Nevertheless, we see no reason to dismiss, in principle, the possibility that cross-linguistic contact might give rise to de-merger. Finally, we note in closing that the present analysis provides a compelling case for the utility of legacy data for tracing language change.

References

Andreasson, Helene N. and Kathrine Asla Østby. 2014. Une étude inter-variété des voyelles moyennes antérieures arrondies en français: une approche phonético-phonologique. In *La phonologie du français: Norms, périphéries, modélisation* ed. by Jacques Durand, Gjert Kristofferson & Bernard Laks, 35–81. Paris: Presses Universitaires de Paris Ouest.

Bazylko, Sławomir. 1981. "Le statut de [ə] dans le système phonématique du français contemporain et quelques questions connexes." *La linguistique* 17 (1): 91–101.

Boersma, Paul, and David Weenink. 2013. Praat: doing phonetics by computer [Computer program]. Version 5.3.52, accessed December 14, 2013, http://www.praat.org/

Bounds, Paulina, Naomi Palosaari, and William A. Kretzschmar, Jr. 2011. Issues in Using Legacy Data." In *Sociophonetics: A Student's Guide*, ed. by Marianna Di Paolo and Malcah Yaeger-Dror, 46–57. London: Routledge.

Bullock, Barbara E., and Chip Gerfen. 2004a. "Phonological Convergence in a Contracting Language Variety." *Bilingualism: Language and Cognition* 7 (2): 95–104. doi: 10.1017/S1366728904001452

Bullock, Barbara E., and Chip Gerfen. 2004b. "Frenchville French: A Case Study in Phonological Attrition." *International Journal of Bilingualism* 8 (3): 303–320. doi: 10.1177/13670069040080030801

Bullock, Barbara E., and Chip Gerfen. 2005. "The Preservation of Schwa in the Converging Phonological System of Frenchville (PA) French." *Bilingualism: Language and Cognition* 8 (2): 117–130. doi: 10.1017/S1366728905002178

Bürki, Audrey, Fougeron, Cécile, Gendrot, Cedric, and Ulrich H. Frauenfelder. 2011. "Phonetic Reduction Versus Phonological Deletion of French Schwa: Some Methodological Issues." *Journal of Phonetics* 39: 279–288. doi: 10.1016/j.wocn.2010.07.003

Colantoni, Laura, and Jeffrey Steele. 2011. "Synchronic Evidence of a Diachronic Change: Voicing and Duration in French and Spanish Stop–Liquid Clusters." *The Canadian Journal of Linguistics / La revue canadienne de linguistique* 56 (2): 147–177. doi: 10.1353/cjl.2011.0016

Coveney, Aidan. 2001. *The Sounds of Contemporary French: Articulation and Diversity*. Exeter: Elm Bank Publications.

Delattre, Pierre. 1966. *Studies in French and Comparative Phonetics*. The Hague: Mouton.

Di Paolo, Marianna, and Alice Faber. 1990. "Phonation Differences and the Phonetic Content of the Tense-Lax Contrast in Utah English." *Language Variation and Change* 2 (2): 155–204. doi: 10.1017/S0954394500000326

Faber, Alice, and Marianna Di Paolo. 1995. "The Discriminability of Nearly Merged Sounds". *Language Variation and Change* 7: 35–78. doi:10.1017/S0954394500000892

Fougeron, Cécile, Cedric Gendrot, and Audrey Bürki. 2007. "On the Phonetic Identity of French Schwa Compared to /ø/ and /œ/." Paper presented at the International Congress of Phonetic Sciences, Saarbrücken, Germany, August 6–10, 2007.

Hall, Kathleen C., and Elizabeth V. Hume. 2013. "Perceptual Confusability of French Vowels." *Proceedings of Meetings on Acoustics* 19: 060113. doi:10.1121/1.4800615

Hansen, Anita Berit. 1994. "Étude du E caduc: stabilisation en cours et variations lexicales." *Journal of French Language Studies* 4: 25–54. doi:10.1017/S0959269500001964

Hansen, Anita Berit. 2012. "A Study of Young Parisian Speech: Some Trends in Pronunciation." In *Phonological Variation in French: Illustrations from Three Continents*, ed. by Randall Scott Gess, Chantal Lyche, and Trudel Meisenburg, 151–172. Amsterdam: John Benjamins. doi:10.1075/silv.11.09han

Hickey, Raymond. 2004. "Mergers, Near-Mergers and Phonological Interpretation." In *New Perspectives on English Historical Linguistics*, ed. by Christian J. Kay, Carole Hough, and Irené Wotherspoon, 125–137. Amsterdam: John Benjamins. doi:10.1075/cilt.252.10hic

Janson, Tore, and Richard Schulman. 1983. "Non-distinctive Features and Their Use." *Journal of Linguistics* 19 (2): 321–336. doi:10.1017/S0022226700007763

Labov, William. 1994. *Principles of Linguistic Change. Volume 1: Internal Factors*. Oxford: Blackwell.

Labov, William. 2007. "Transmission and Diffusion." *Language* 83 (2): 344–387. doi:10.1353/lan.2007.0082

Labov, William, Mark Karen, and Corey Miller. 1991. "Near-mergers and the Suspension of Phonemic Contrast." *Language Variation and Change* 3 (1): 33–74. doi:10.1017/S0954394500000442

Labov, William, Malcah Yaeger, and Richard Steiner. 1972. *A Quantitative Study of Sound Change in Progress* (Vol. 1). Philadelphia: The U.S. Regional Survey.

Léon, Pierre. 2007. *Phonétisme et pronunciations du français*, 5e édition. Paris: Armand Colin.

Malécot, André. 1974. "The Frequency of Occurrence of Phonemes and Consonant Clusters in French." *Phonetics* 29: 158–170. doi:10.1159/000259468

Malécot, André, and Gérard Chollet. 1977. "The Acoustic Status of the Mute-e in French". *Phonetica* 34 (1): 19–30. doi:10.1159/000259866

Manaster Ramer, Alexis. 1996. "A Letter from an Incompletely Neutral Phonologist." *Journal of Phonetics* 24 (4): 477–489. doi:10.1006/jpho.1996.0026

Martinet, André. 1972. "La nature phonologique d'e caduc." In *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*, ed. by Albert Valdman, 393–399. The Hague: Mouton.

Mielke, Jeff. 2013. "Ultrasound and Corpus Study of a Change from Below: Vowel Rhoticity in Canadian French." *University of Pennsylvania Working Papers in Linguistics* 19 (2): 141–150.

Morin, Yves-Charles. 1978. "The Status of Mute 'e'." *Studies in French Linguistics* 1 (2): 79–140.

Morin, Yves-Charles. 2000. "Le français de référence et les normes de pronunciation." *Cahiers de l'Institut de Linguistique de Louvain* 26 (1): 91–135. doi:10.2143/CILL.26.1.503937

Mougeon, Françoise, and Dorin Uritescu. 2006. "Les derniers Français de Frenchville." *La Linguistique* 42 (1): 115–128. doi:10.3917/ling.421.0115

Tranel, Bernard. 1987. "French Schwa and Nonlinear Phonology." *Linguistics* 25: 845–866. doi:10.1515/ling.1987.25.5.845

Valdman, Albert. 1979. "Créolisation, français populaire et le parler des isolats francophones d'Amérique du Nord." In *Le français hors de France*, ed. by Albert Valdman, 181–197. Paris: Champion.

Verluyten, S. Paul. 1988. *La phonologie du schwa français*. Amsterdam & Philadelphia: John Benjamins. doi: 10.1075/lis.16

Walker, Douglas C. 1993. "Schwa and /œ/ in French." *Canadian Journal of Linguistics* 38 (1): 43–64.

Walter, Henriette. 1982. *Enquête phonologique du français*. Paris: Presses universitaires de France.

Wassink, Alicia Beckford. 2006. "A Geometric Representation of Spectral and Temporal Vowel Features: Quantification of Vowel Overlap in Three Varieties." *Journal of the Acoustical Society of America* 119 (4): 2334–2350. doi: 10.1121/1.2168414

Yu, Alan C. L. 2007. "Understanding Near Mergers: The Case of Morphological Tone in Cantonese." *Phonology* 24 (1): 187–214. doi: 10.1017/S0952675707001157

CHAPTER 12

The processing of intrasentential anaphoric subject pronouns in L2 Spanish

Juan P. Comínguez,* Nuria Sagarra,* Aurora Bel**
and Estela García-Alcaraz**
* Rutgers University / ** Universitat Pompeu Fabra

This paper examines whether L2 learners of Spanish can attain native-like interpretive and parsing patterns in intrasentential subject anaphor resolution, and the extent to which these patterns are modulated by L1 transfer. Adult advanced L2 learners of Spanish with different L1s (Moroccan Arabic, a null-subject language, and English, a non-null subject language) completed an experiment utilizing online (the moving window paradigm) and offline (answers to comprehension questions) techniques. Both Arabic and English learners of Spanish performed in a native-like way at the interpretive and parsing levels. These patterns suggest that L1 transfer plays no crucial role when interpreting and processing intrasentential anaphoric subjects at least at the advanced level.

Keywords: null-subject languages, subject pronouns, intrasentential subject anaphora, language parsing, L2 Spanish

1. Introduction

A well-studied domain of the syntax-pragmatics interface phenomena in adult L2 acquisition is the discourse properties licensing null and overt subject pronouns, as well as their antecedent biases (cf. Montrul 2004; Sorace 2011). In addition to acquiring the morphosyntactic properties allowing the existence of null subjects in null-subject languages, adult L2 learners have to master pragmatic properties and anaphor resolution biases constraining the interpretation of pronominal subjects. For instance, in the *topichood* dimension of information structure, null pronominal subjects realize *discourse topics* and obtain their reference from preceding subjects, while their overt counterparts function as topic-shifts and get their interpretation from non-subjects (Gutiérrez-Bravo 2008, a.o.).

DOI 10.1075/rllt.11.12com

Existing evidence on late L2 acquisition of those pragmatic constraints and the anaphor resolution biases is inconclusive. At the intrasentential level, several studies have shown persistent difficulties consisting of L2 learners overextending the scope of overt subject pronouns over their null counterparts (e.g. Keating, VanPatten, & Jegerski 2011; Sorace & Filiaci 2006, a.o.). This overextension consists of utilizing overt subject pronouns to realize *discourse topics* and solving the anaphoric dependency with preceding subjects. Nevertheless, some studies have reported successful cases of L2 acquisition of these constraints and biases (e.g. Kras 2008; Bel & García-Alcaraz 2015, a.o.), but from their experimental manipulation, it is not possible to determine whether proficiency level or positive L1 transfer underlies this convergent outcome.

This study investigates the extent to which L1 transfer modulates the online and offline intrasentential anaphor resolution patterns in L2 Spanish. Data elicited via online (moving-window paradigm) and offline (answers to comprehension questions) techniques provide us with a general picture of the parsing strategies implemented by L2 learners in real time, as well as the interpretive outcome of the intrasentential subject anaphor resolution process.

2. Morphosyntactic and anaphoric properties of null and overt subject pronouns in null-subject languages

Whereas some natural languages like English require referential and expletive sentence subject arguments to be overt (1), other languages like Spanish allow the existence of referential null subjects (2a), and require expletive subjects to be null (2b).

(1) a. **(Pedro) / *(He) ate potatoes.*
b. **(It) rains.*

(2) a. *(Pedro) / (Él) comió patatas.*
'Pedro / He ate potatoes.'
b. *(*Ello) llueve.*
'It rains.'

The differences between languages like Spanish and English concerning null subjects were captured by proposing the existence of the Null Subject Parameter (Chomsky 1981, 1982; Rizzi 1982, 1986, 1997, a.o.). In Rizzi's (1986, 1997) analysis, a null subject is analyzed as a phonetically empty category called *pro*. According to this proposal, there are two universal linguistic principles subject to parameterization: (1) *pro* must be licensed and (2) identified by the governing head Infl°. Thus, the differences between (1) and (2) follow from the parameterization of those two

linguistic principles, whereas Spanish has an Infl^{o} that licenses and identifies *pro*, English does not.

In addition to these morphosyntactic properties, null-subject languages of the Spanish type (e.g. Italian, Catalan, etc.) exhibit a division of labor between null and overt subject pronouns with regard to how they obtain their interpretation due to their anaphoric nature. Carminati (2002) explored the anaphor resolution process between a subject pronoun and its antecedent in Italian. In a series of experiments employing online and offline methods, native speakers of Italian displayed an online processing bias strategy according to which null subject pronouns (henceforth, NSPs) in embedded clauses like in (3) were preferably interpreted as being coreferential with matrix subject DPs.

(3) *Marta$_i$ scriveva frequentementa a Piera quando ø$_i$ era negli Stati Uniti.*
'Marta used to write frequently to Piera when she was in the United States.'
(Carminati 2002: 83)

By contrast, overt subject pronouns (henceforth, OSPs) in embedded clauses like in (4) were preferably interpreted as being coreferential with matrix object DPs.

(4) *Marta scriveva frequentementa a Piera$_i$ quando lei$_i$ era negli Stati Uniti.*
'Marta used to write frequently to Piera when she was in the United States.'
(Carminati 2002: 83)

Carminati (2002) proposed the Position of the Antecedent Hypothesis (PAH) in order to account for these patterns. The PAH predicts that in null-subject languages NSPs prefer an antecedent in [Spec, IP], whereas OSPs prefer an antecedent which is not in [Spec, IP]. This processing bias is the result of a pragmatic interpretive preference based on structural information that constrains the sentence parser when resolving the anaphoric dependencies between subject pronouns and their antecedents in null-subject languages.

Nevertheless, the extent to which the PAH is operative in other null-subject languages, as proposed by Carminati (2002), is far from being clear. In a series of online experiments, Filiaci, Sorace, and Carreiras (2013) found that Peninsular Spanish only exhibited a strong processing and interpretive bias with respect to NSPs, whereas Italian showed biases in relation with NSPs as well as OSPs. Specifically, data from Peninsular Spanish speakers reveal that whereas NSPs were consistently assigned subject DPs as their antecedents, OSPs solved their antecedent dependency with either subject DPs or constituents in non-subject positions. However, Keating, Jegerski, and VanPatten (2015) conducted a similar online experiment on the operativeness of the PAH in Mexican Spanish, and discovered that this variety of Spanish has an Italian-like online preference with OSPs. Along the lines of the patterns reported by Filiaci et al. (2013), Bel and García-Alcaraz

(2015) found a similar interpretive bias of the Peninsular Spanish type in Moroccan Arabic (henceforth, MA). The results of an offline study showed that NSPs were preferably assigned subject DPs as their antecedents, but OSPs could obtain their antecedents from either subject DPs or non-subject DPs. Taken together, the findings from Filiaci et al. (2013) and Bel and García-Alcaraz (2015) suggest that not all the null-subject languages follow the PAH as defined by Carminati (2002) for Italian. Instead, other null-subject languages like Peninsular Spanish and MA only show a bias with null subject pronouns.

In non-null-subject languages, there are also anaphoric processing biases involving subject pronouns. For example, English exhibits the existence of a subject antecedent bias (Crawley, Stevenson, & Kleinman 1990; Frederiksen 1981). According to this bias, the sentence parser prefers to assign subject DPs as antecedents for subject pronouns. Nevertheless, factors like implicit causality (Arnold 2001; Garvey & Caramazza 1974, a.o.), plausibility (Sanford & Garrod 1981; Stevenson & Vitkovitch 1986, a.o.), mention order (Gernsbacher 1989; Gernsbacher & Hargreaves 1988, 1992), parallelism (Sheldon 1974; Stevenson, Crawley, & Kleinman 1994, a.o.), and phonetic cues (Sheldon 1974, a.o.) also modulate anaphor resolution. In sum, in non-null-subject languages like English there is a set of factors modulating subject anaphor resolution that supplement the subject parsing bias.

3. Adult L2 acquisition of pragmatic and semantic properties of subject pronouns in null-subject languages

The adult L2 acquisition of the morphosyntactic properties characterizing null-subject languages of the Spanish type is a well-studied phenomenon in the field of second language acquisition (Al-Kasey & Pérez-Leroux 1998; Phinney 1987; White 1985, a.o.). The main generalization resulting from these findings is that both expletive and referential null subjects are acquired early and fast by L2 learners, regardless of the *pro*-drop or non-*pro*-drop nature of their L1s. In contrast, considerable difficulties have been found in the non-native mastering of the subject anaphor resolution patterns (Al-Kasey & Pérez-Leroux 1998; Bini 1993; Margaza & Bel 2006; Rothman 2007, 2009, a.o.).

At the intrasentential level, a pattern characterized by the non-native overextension of the scope of OSPs over NSPs has been found. Sorace and Filiaci (2006) tested the operativeness of Carminati's (2002) PAH in L1 English near-natives of L2 Italian, and found a non-native-like pattern consisting of choosing matrix subject DPs as the antecedents of OSPs with either preposed or postposed temporal embedded clauses. By contrast, near-natives performed in a native-like way in the antecedent assignment of NSPs (for similar results, see Belletti, Bennati & Sorace

2007). However, Kras (2008) found that L1 Croatian near-natives of L2 Italian performed in a native-like way with respect to the anaphor resolution patterns constrained by Carminati's (2002) PAH. In a task very similar to the one utilized by Sorace and Filiaci (2006), non-natives showed native-like patterns concerning the interpretation of NSPs and OSPs. From these findings, Kras concluded that the L1 of the L2 learners might modulate the acquisition of the parsing strategies and the pragmatic properties licensing null and overt subject pronouns. This is because Croatian is a null-subject language in which the PAH is operative in the same way as it is in Italian.

Further offline studies with L1 English L2 Spanish learners of lower proficiency found subject anaphor resolution patterns to be non-native-like. The pattern of results revealed that OSPs and NSPs were in free variation with respect to subject anaphor resolution patterns (Keating et al. 2011; Jegerski, VanPatten & Keating 2011). Nevertheless, in Jegerski et al.'s (2011) study, advanced learners exhibited an increasing sensitivity to the PAH constraints consisting of selecting matrix subjects as the antecedents for NSPs. Taken together, these data indicate a tendency for narrowing the subject anaphor resolution patterns from free variation to only over-extending overt subject pronouns. More recently, Bel and García-Alcaraz (2015) investigated the offline subject backwards and forward anaphor resolution preferences in L1 MA upper-intermediate learners of L2 Spanish. Their findings showed that L2 learners performed in a native-like fashion with both null and overt subject backwards anaphors, but not with their forward counterparts.

Summarizing, such opposite outcomes are likely related to different factors (e.g. L1 transfer, L2 proficiency, morphosyntactic and semantic biases of pronouns, etc.), and it is not easy to determine which factors modulate them. The existing evidence on the operativeness of the PAH in developing and steady-state L2 systems is inconclusive. For these reasons, this study seeks to contribute to this debate in the field of second language acquisition by examining the role of L1 (Moroccan Arabic vs. English) in the interpretation and processing patterns of intrasentential subject anaphors in L2 Spanish.

4. Research questions and hypotheses

The research questions of the current study are the following:

RQ1: Do native and non-native speakers show the same patterns for null and overt antecedent biases in terms of intrasentential anaphor interpretation and processing?

Prediction: If the PAH is operative, both in native and non-native systems, we expect NSPs to be clearly biased towards subject antecedents in contrast with OSPs.

RQ2: Does L1 transfer have an effect on subject anaphor resolution in L2 Spanish?

Prediction: We predict that, due to linguistic proximity concerning anaphor resolution patterns, MA learners will perform in a more native-like way than English learners.

5. Experiment

5.1 Participants

Participants consisted of 32 Spanish native speakers (data collected in Spain, where they were born and raised), 44 English advanced learners of Spanish (data collected in the U.S.), and 44 Arabic advanced learners of Spanish (data collected in Morocco). The three groups were similar in education (at least high school degree), handedness (right), age (18–37 years old), working memory (one-way ANOVA with Group as independent variable: $F(2,107) = 1.890$, $p = .156$), and subject bias preferences (see accuracy results). In addition, the two learner groups were homogeneous in age of onset (all began learning Spanish post-puberty), Spanish proficiency (independent samples *t*-test on DELE test scores: $t(84) = 1.26$, $p = .211$), lexical knowledge (independent samples *t*-test on vocabulary test scores: $t(68) = 1.37$, $p = .175$), grammatical knowledge (independent samples *t*-test on grammar test scores: $t(68) = 1.50$, $p = .139$), and study abroad experience (0–3 months). Finally, the English learners had no knowledge of other languages, and the Arabic learners knew some Modern Standard Arabic and French.

5.2 Materials and procedure

Participants completed six tests: a language background questionnaire (5 minutes), a Spanish proficiency test (25 minutes), a working memory test (15 minutes), a non-cumulative word-by-word self-paced reading task (20 minutes), a vocabulary test (10 minutes), and a grammar test (3 minutes).

5.2.1 *Language background questionnaire*

This test asked questions in the participants' L1 about age of acquisition, previous exposure to Spanish, contact with Spanish outside the course, and knowledge of other languages.

5.2.2 *Spanish proficiency test (learners only)*

This test was adapted from the *Diploma de Español como Lengua Extranjera,* "Diploma of Spanish as a Foreign Language" exams, which measure the degree of fluency in the Spanish Language, and are issued and recognized by the Ministry of Education, Culture and Sport of Spain. The DELE exams are prepared by the University of Salamanca and distributed by the Instituto Cervantes. The adapted version used in this study had a 36-item multiple-choice grammar section and a 20-item multiple-choice reading comprehension section. Adapted versions of DELE exams are widely used in L2 acquisition and sentence processing studies (e.g. Slabakova, Kempchinsky, & Rothman 2012; Keating et al. 2015).

5.2.3 *Working memory test*

A letter-number sequencing test assessed working memory by showing participants sets of two to nine letters and numbers in a random non-cumulative manner one at a time (e.g. 7-J-M-3), and by asking them to recall first the numbers in ascending order and then the letters in alphabetical order (e.g. 37JM). This test was adapted from a subtest from the revised version of the Wechsler Adult Intelligence Scale test (WAIS) (Wechsler 1997), and it is widely used in L2 studies examining L1 transfer effects due to its non-linguistic nature (e.g. Sagarra and Ellis 2013).

5.2.4 *Self-paced reading task*

In this task, participants read sentences on a computer screen, word-by-word, in a left-to-right display. For each sentence, participants looked at a 500-ms fixation sign (+) and saw dashes indicating the words that would appear and how many letters each word would have, to make reading more natural. Then they pressed the spacebar key to make the first word appear. Each subsequent time they pressed the spacebar key, the next word appeared and the prior word disappeared. After each sentence, participants answered a question by pressing a "yes" or a "no" button, and sentences were balanced for yes and no answers. In total, participants read 111 sentences: 3 practice, 84 filler, and 24 experimental (6 per condition). The presentation order changed for each learner, following a Latin square design with 6 blocks of sentences, one sample of each condition per block, and blocks scrambled in a pseudo-randomized manner.

Each experimental sentence had 4 conditions: two with S(ubject) antecedent – (5) null pronoun, (6) overt pronoun –, and two with O(bject) antecedent – (7) null pronoun, (8) overt pronoun.

(5) *El músico$_i$ saluda al bombero$_j$ mientras Ø$_i$ lleva un violín en la mochila*
'The musician greets the fireman as he carries a violin in his backpack.'

(6) *El músico$_i$ saluda al bombero$_j$ mientras él$_i$ lleva un violín en la mochila*
'The musician greets the fireman as he carries a violin in his backpack.'

(7) *El músico$_i$ saluda al bombero$_j$ mientras Ø$_j$ lleva un casco en la mochila*
'The musician greets the fireman as he carries a helmet in his backpack.'

(8) *El músico$_i$ saluda al bombero$_j$ mientras él$_j$ lleva un casco en la mochila*
'The musician greets the fireman as he carries a helmet in his backpack.'

Experimental sentences were 12–13 words long, had 2- to 4-syllable human S and O with transparent gender and singular number in the main clause, and 2- to 3-syllable inanimate O in the subordinate clause. The sentences were balanced for pronoun type (50% null, 25% masculine overt, 25% feminine overt) and antecedent type (25% masculine S, 25% feminine S, 25% masculine O, 25% feminine O). Matrix verbs had neutral implicit causality for interpreting the pronoun as co-referent with either S or O (bias range towards S = 47–59%, based on Goikoetxea, Pascual and Acha (2008), and each verb was used twice (masculine S sentence, feminine S sentence). Finally, 50% of the sentences had questions (e.g. *¿El músico lleva un violín?* 'The musician carries a violin?') assessing interpretation (25% in S antecedent conditions, 25% in O antecedent conditions, like the example sentence above), and 50% assessing S/O preference (25% in O antecedent conditions, 25% in S antecedent conditions, like the example sentence above).

5.2.5 *Vocabulary test (learners only)*

This test evaluated the learners' familiarity with the target nouns (matrix S, matrix O, subordinated O) by asking them to match the Spanish nouns with their corresponding English or Moroccan Arabic translation. All scored above 75%, with no statistical differences between the two learner groups, $t(68) = 1.37$, $p = .175$.

5.2.6 *Grammar test (learners only)*

This test evaluated the learners' knowledge of the verb forms used in experimental and filler sentences (regular verbs, 3rd person singular/plural, present/preterit) by asking them to match Spanish conjugated verbs with their corresponding English or Moroccan Arabic translation. All scored above 86%, with no statistical differences between the two learner groups, $t(68) = 1.50$, $p = .139$.

5.3 Scoring

The scoring of the Spanish proficiency test, the vocabulary test, and the questions (both interpretation and S/O preference) of the self-paced reading task was 1 point for correct answers and 0 for incorrect ones. To score 1 point per set in the working memory test, participants had to recall the letters and numbers in the correct order.

Finally, online processing of the sentences of the self-paced reading task was measured with residual reading times (RTs) at the verb (V), the object (NP), and the PP (*en la mochila* 'in the backpack'). Residual RTs were calculated to eliminate possible biases due to individual reading speeds or sentence length, following Ferreira and Clifton (1986), and were later converted to *T* scores to allow a Gamma distribution, which provides a better fit to this type of data than normal distribution (Anderson, Verkuilen, & Johnson 2013).

6. Results

Five Generalized Linear Mixed Models (GLMMs) were conducted on separate data points (not aggregate means), one per dependent variable: residual RTs at the V, object NP, and PP, accuracy at selecting the antecedent (subject bias), and accuracy at interpreting the sentences. All GLMMs had Pronoun (null, overt), Antecedent (S, O), and Group (Spanish natives, English learners, Arabic learners), and all their possible interactions, as fixed factors, and Subject as a random factor. The first three GLMMs followed a gamma distribution (log link) and the other two a binominal distribution (logit link). Alpha level was .05, pairwise comparisons were calculated with Bonferroni post-hoc *t*-tests, and only significant effects and interactions were reported. Finally, a conservative approach was adopted when building the GLMMs, by selecting the Satterthwaite approximation for *df* to handle small groups and possible abnormally distributed data, and the robust estimation for the test of fixed effects and coefficients to control potential violations for model assumptions.

6.1 RTs at the V

Descriptive statistics can be found in Table 1.

Table 1. Descriptive statistics of means in milliseconds at the V

Group	Null pronoun				Overt pronoun			
	S antecedent		O antecedent		S antecedent		O antecedent	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Spanish native speakers	471.35	48.57	466.57	49.12	468.27	63.22	466.80	50.23
Arabic learners	471.16	74.06	471.15	75.81	473.86	73.42	466.54	65.82
English learners	464.65	62.51	467.68	66.27	461.68	61.05	462.28	73.79

All fixed factors were found to be non-significant.

6.2 RTs at the object NP

Descriptive statistics can be found in Table 2.

Table 2. Descriptive statistics of means in milliseconds at the object NP

Group	Null pronoun				Overt pronoun			
	S antecedent		O antecedent		S antecedent		O antecedent	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Spanish native speakers	1,000.05	147.33	1,014.54	160.97	1,029.84	170.05	1,024.28	154.04
Arabic learners	962.00	165.09	960.30	153.78	973.34	172.43	971.29	187.74
English learners	975.02	155.31	979.73	163.05	974.42	142.36	970.21	147.20

A main effect of Group was found to be significant, $F(2, 117) = 10.977$, $p = .000$. Pairwise contrasts suggest that Spanish natives were slower than both English learners ($p = .001$) and Arabic learners ($p = .000$), with no statistical difference between the two learner groups ($p = .420$).

6.3 RTs at the PP

Descriptive statistics can be found in Table 3.

Table 3. Descriptive statistics of means in milliseconds at the PP

Group	Null pronoun				Overt pronoun			
	S antecedent		O antecedent		S antecedent		O antecedent	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Spanish native speakers	1,759.12	327.33	1,811.90	370.08	1,713.08	306.16	1,697.80	269.73
Arabic learners	1,498.68	296.57	1,509.68	315.99	1,503.20	323.00	1,417.54	214.76
English learners	1,496.57	263.20	1,545.83	288.20	1,536.37	286.84	1,502.94	287.22

Two main effects and two paired interactions were found to be significant. There was a main effect of Group, $F(2, 116) = 48.754$, $p = .000$, such that Spanish natives were slower than both English learners ($p = .000$) and Arabic learners ($p = .000$), with no statistical difference between the two learner groups ($p = .125$). In addition, there was a main effect of Pronoun, $F(1, 2602) = 13.839$, $p = .000$, such that PP regions of clauses with null subject pronouns were read slower than PP regions of clauses with overt subject pronouns ($p = .000$). The first significant interaction was Pronoun ×

Antecedent, $F(1, 2602) = 14.012$, $p = .000$, suggesting the following: first, PP regions of clauses with null subject pronouns were read slower than PP regions of clauses with overt subject pronouns when the antecedent was the object ($p = .000$), but not when the antecedent was the subject ($p = .987$); secondly, PP regions of clauses with null subjects were read slower with object antecedents than with subject antecedents ($p = .025$), but the opposite obtained with PP regions of clauses with overt subject pronouns, i.e. there were slower reading times in PP regions of clauses with subject antecedents than with object antecedents ($p = .002$). The second significant interaction was Group × Pronoun, $F(2, 2602) = 3.495$, $p = .031$, which suggests that Spanish natives were slower with both null and overt subject pronouns than both English learners ($p = .000$) and Arabic learners ($p = .000$). Also, there was a significant difference between both learner groups, which consisted of English learners being slower than Arabic learners only with overt subject pronouns ($p = .036$). Moreover, it is worth mentioning the marginally significant interaction of Group × Antecedent, $F(2, 2601) = 2.673$, $p = .069$, which sheds light on the main effect of Pronoun in the Group × Pronoun interaction described above. Specifically, Spanish natives were slower with subject and object antecedents than both English learners ($p = .000$) and Arabic learners ($p = .000$). Along these lines, there was a significant difference between both learner groups, which consisted of English learners being slower than Arabic learners only with object antecedents ($p = .028$).

6.4 Accuracy – subject bias

Descriptive statistics can be found in Table 4.

Table 4. Descriptive statistics for accuracy data (subject bias) in percentages

Group	Null pronoun				Overt pronoun			
	S antecedent		O antecedent		S antecedent		O antecedent	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Spanish native speakers	70.59	45.83	58.89	49.48	50.59	50.29	32.22	46.99
Arabic learners	65.96	47.55	50.00	50.17	47.92	50.13	43.33	49.72
English learners	70.49	45.80	70.45	45.80	69.60	46.18	59.09	49.35

The three main effects were found to be significant, but no interaction was significant. For Group, $F(2,46) = 6.453$, $p = .003$, English learners were more subject-biased than both Arabic learners ($p = .004$), and Spanish natives ($p = .018$), with no difference between the last two ($p = .833$). For Pronoun, $F(1, 1434) = 27.624$, $p = .000$, all groups were more subject-biased in sentences with null than with overt pronouns ($p = .000$). Finally, for Antecedent, $F(1, 1434) = 14.845$, $p = .000$,

all groups were more subject-biased in sentences with a subject antecedent than with an object antecedent (p = .000).

6.5 Accuracy – sentence interpretation

Descriptive statistics can be found in Table 5.

Table 5. Descriptive statistics for accuracy data (interpretation) in percentages

Group	Null pronoun				Overt pronoun			
	S antecedent		O antecedent		S antecedent		O antecedent	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Spanish native speakers	99.02	09.90	94.79	22.34	98.04	13.93	96.91	17.40
Arabic learners	87.72	32.97	84.26	36.59	89.47	30.82	86.36	34.47
English learners	78.79	41.04	87.40	33.31	84.85	35.99	85.48	35.37

Only the main effect of Group was found to be significant, $F(2, 69) = 5.918$, p = .004, such that Spanish native speakers understood the meaning of the sentences better than both English learners (p = .013) and Arabic learners (p = .021), with no difference between the two learner groups (p = .778).

7. Discussion

This study examined the online intra-sentential interpretive patterns of null and overt subject pronouns in native and non-native Spanish, and the extent to which those patterns were modulated by L1 transfer. In particular, the two research questions focused on: (1) whether native and non-native speakers of Spanish showed the same interpretive and parsing patterns in intrasentential subject anaphor resolution, and (2) the role of L1 transfer in this process.

With regard to the first research question, results on interpretive patterns revealed that there is a strong subject bias according to which NSPs obtain their antecedent from matrix subjects, as predicted by Carminati's (2002) PAH. Importantly, this pattern is attested in all groups, which reveals the existence of native-like intrasentential subject anaphor resolution. Along these lines, the results on the parsing patterns point in the same direction, since all the groups equally complied with the PAH in real time. Specifically, both natives and non-natives exhibited faster RTs with subject antecedents and NSPs, as well as with object antecedents and OSPs. That is, the PP region was read faster when a NSP had a matrix subject as its antecedent. In contrast, the same region was read slower when a NSP had a matrix

object as its antecedent. Moreover, the PP region with OSPs was read faster when an OSP had the matrix object as its antecedent than when its antecedent was the matrix subject. Taken together, these processing patterns evince the online operativeness of Carminati's PAH in both native and non-native speakers of Spanish.

These online and offline findings are consistent with previous offline studies reporting convergence at the interpretive level with regard to intrasentential subject anaphor resolution (Kras 2008; Bel & García-Alcaraz 2015), which suggests that linguistic phenomena involving the syntax-pragmatics interface do not constitute insurmountable difficulties for L2 learners, as advanced by Sorace and Filiaci (2006), and Belletti et al. (2007), among others. Nevertheless, it is worth mentioning that the offline experimental tasks in studies such as those by Sorace and Filiaci (2006) and Belletti et al. (2007) were more complex. Their complexity consisted in having three possible antecedents for intrasentential subject anaphors: the matrix subject, the matrix object, and an extralinguistic antecedent not mentioned in the matrix clause. Hence, the role of task complexity in intrasentential subject anaphor resolution requires being experimentally addressed in a direct fashion before reaching further conclusions about the difficulties posed by this syntax-pragmatics interface phenomenon to L2 learners.

The existing evidence (Jegerski et al. 2011; Keating et al. 2011, a.o.) reveals that L2 learners of null-subject languages undergo a developmental path consisting of instantiating null subjects in their interlanguages, and further adjusting the intrasentential subject anaphor resolution patterns. These patterns begin showing free variation between NSPs and OSPs with regard to their antecedents. Then, they exhibit an overextension of the interpretive options only for OSPs, and finally display a convergence with the native preferences, as shown by the findings reported in this paper. There is, however, an issue that deserves further consideration, namely, the fact that advanced L2 learners in this study and those in previous studies by Jegerski et al. (2011) and Keating et al. (2011) resolve intrasentential subject anaphors in different ways in spite of their similar levels of proficiency. Specifically, whereas advanced L2 learners in this experiment displayed a native-like intrasentential anaphor resolution, that was not the case with advanced L2 learners in the aforementioned studies. The differences in the experimental methods and design may provide us with a possible way of accounting for these asymmetries. First, the offline methodology utilized in the previous studies may have obscured the actual employment of a parsing bias such as the PAH. As Keating et al. (2011) observe, untimed offline judgments are highly influenced by non-linguistic phenomena such as introspection, which can lead to other factors determining the interpretation of the anaphors. Along these lines, those studies did not control for verbal implicit causality (see Section 2), which can be argued to operate as an underlying

confounding factor resulting in unexpected interpretation patterns in L1 English learners of L2 Spanish.

With regard to the second research question, results show some surface L1 transfer effects, in spite of the attested native-like interpretive and processing patterns in both learner groups. In particular, it is noteworthy that English learners displayed a stronger subject bias at the interpretive level than both Spanish natives, and Arabic learners. This offline phenomenon in English learners, together with the fact that they were also slower than Arabic learners with OSPs and object antecedents, can be taken as a surface effect related to the influence of the English subject parsing bias (see Section 2). However, this surface effect does not prevent English learners from converging with Arabic learners and Spanish natives at both the interpretive and parsing levels. This convergence demonstrates the instantiation of null subjects in the interlanguage of these advanced L2 learners. Otherwise, the operativeness of a native-like PAH would not take place when L2 learners solve the anaphor-antecedent dependency.

There are two processing effects found in this study that are worth mentioning. The first effect concerns the pace at which the PP region with NSPs and OSPs was read. Specifically, clauses with OSPs displayed faster reading times, whereas sentences with NSPs exhibited processing penalty effects, even in contexts that favored a subject antecedent. These results suggest that NSPs are, in general, more costly to process than OSPs, which is in line with findings from other data sources. Yamashita, Stowe and Nakayama (1993) and Sagarra and Ellis (2013) provide experimental online evidence that empty categories and inflectional morphology are more costly to process than overt categories and free morphemes. It can be argued that real interpretive biases could be hidden by the excessive demands that NSPs by themselves pose, thus, creating a processing load; presumably, the sentence processor selects a potential antecedent relying mainly on parsing principles such as Recency, while the configurational and pragmatic properties that shape the understanding of anaphor resolution (e.g. givenness and accessibility, as in Ariel's (1990) Accessibility Theory) *force* speakers to select an element in subject position for a NSP.

The second effect concerns the fact that Spanish natives were slower than all the L2 learners at the object NP and PP regions. The general difference in the RTs between natives and non-natives can be argued to result from the fact that wrap-up effects are stronger in the former than in the latter. In these regions of the clause in which disambiguation takes place, the processes of cognitive and semantic comprehension lead to the full integration of different levels of linguistic and non-linguistic knowledge and procedures (Just & Carpenter 1980). It is relevant to remark that wrap-up effects occur at the end of a sentence or clause, and they are not strictly dependent on the last word of the sentence or clause (Frazier 1999). As

native linguistic (e.g. the grammar) and non-linguistic (e.g. the oculomotor system utilized for reading) systems are complete and steady, slower native RTs point to greater levels of processing complexity, since all the factors are simultaneously accessed, evaluated, and coordinated in order to achieve the aforementioned full integration. By contrast, non-native linguistic and non-linguistic systems are still incomplete and unstable, which results in a simplification of the integration process leading to faster RTs. The fact that Spanish natives were significantly more accurate in interpretive questions than both English learners and Arabic learners provides further evidence that Spanish natives were slower than L2 learners in these regions due to this process of full integration. In this respect, there is a trade-off between processing and interpretation in these regions consisting of achieving higher comprehension at the expense of a general reading slowdown. In addition, this full integration process seems to operate in an independent way from anaphor resolution, which arguably involves simpler mechanisms, since the parser essentially takes into account syntactic information to solve the anaphor-antecedent dependency (Carminati 2002). To sum up, these processing patterns are expected, as they are a consequence of wrap-up effects in these clausal regions in which the processes of cognitive and semantic comprehension take place.

8. Conclusions

The current study provides evidence that pronominal anaphor resolution in native and non-native Spanish is constrained by the PAH. NSPs are shown to tend to co-refer with preceding matrix subjects, whereas OSPs are not as strongly subject-biased. These results are harmonious with Carminati's (2002) PAH, at least for backwards anaphors, the type of anaphor assessed in this experiment. Future research should investigate the processing of forward anaphors in intrasentential contexts in order to ascertain if the same trends are revealed. As for the role of L1 transfer, this study suggests that the L1 causes some surface effects which do not prevent learners from attaining native-like intrasentential subject anaphor resolution in the L2, at least at an advanced level of proficiency. This is an issue that deserves further research in which lower levels of proficiency are experimentally manipulated. Processing and interpretive data have offered some patterns that provide us with a glimpse of the complexity entailed in assigning antecedents to pronouns in real time; in fact, the performance that native speakers show with interpretive data could be masking the true picture at the cognitive level when processing and interpreting complex phenomena involving the syntax-pragmatics interface. RT data provide us with evidence of how hard the process is, since the native group is the group that spent more time reading clausal regions in which

the pronominal reference must be solved, and the integration of multiple information from different sources takes place in order to obtain an interpretation of the sentence. In this sense, future research should study the role that cognitive factors, such as working memory capacity, play in modulating these processes. Moreover, the current study has also presented evidence that clauses containing NSPs are more costly to process than those with OSPs. Finally, although this experiment has not directly addressed the precise question of development, its results suggest that linguistic phenomena involving the syntax-pragmatics interface do not constitute an insurmountable difficulty for adult second language learners.

Acknowledgements

This research has been founded by a grant from the Spanish Government's Ministerio de Economía y Competitividad (#FFI2012-35058). We would like to thank the *Instituto Cervantes of Marrakech* for its collaborative support.

References

Al-Kasey, Tamara, and Ana T. Pérez-Leroux. 1998. "Second Language Acquisition of Spanish Null Subjects." In *The Generative Study of Second Language Acquisition*, ed. by Suzanne Flynn, Gita Martohardjono, and Wayne O'Neil, 161–185. Mawhah NJ: Lawerence Erlbaum Associates.

Anderson, Carolyn J., Jay Verkuilen, and Timothy Johnson. 2013. *Applied Generalized Linear Mixed Models: Continuous and Discrete Data*. New York: Springer.

Ariel, Mira. 1990. *Accessing Noun Phrase Antecedents*. London/New York: Routledge.

Arnold, Jennifer. 2001. "The Effect of Thematic Roles on Pronoun Use and Frequency of Reference Continuation." *Discourse Processes* 31 (2): 137–162. doi: 10.1207/S15326950DP3102_02

Bel, Aurora, and Estela García-Alcaraz. 2015. "Subjects in the L2 Spanish of Moroccan Arabic Speakers: Evidence from Bilingual and Second Language Learners." In *The Acquisition of Spanish as a Second Language: Data from Understudied Languages Pairings*, ed. by Tiffany Judy, and Silvia Perpiñán, 201–232. Amsterdam: John Benjamins.

Belletti, Adriana, Elisa Bennati, and Antonella Sorace. 2007. "Theoretical and Developmental Issues in the Syntax of Subjects: Evidence from Near-native Italian." *Natural Language and Linguistic Theory* 25: 657–689. doi: 10.1007/s11049-007-9026-9

Bini, Milena. 1993. "La adquisición del italiano: más allá de las propiedades sintácticas del parámetro pro-drop." In *La lingüística y el análisis de los sistemas no nativos*, ed. by Juana M. Liceras, 126–139. Ottawa: Doverhouse.

Carminati, Maria N. 2002. *The Processing of Italian Subject Pronouns*. PhD dissertation, University of Massachusetts.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam. 1982. *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, Mass.: MIT Press.

Crawley, Rosalind A., Rosemary J. Stevenson, and David Kleinman. 1990. "The Use of Heuristic Strategies in the Interpretation of Pronouns." *Journal of Psycholinguistic Research* 194: 245–265. doi:10.1007/BF01077259

Ferreira, Fernanda, and Charles Clifton. 1986. "The Independence of Syntactic Processing." *Journal of Memory and Language* 25: 348–368. doi:10.1016/0749-596X(86)90006-9

Filiaci, Francesca, Antonella Sorace, and Manuel Carreiras. 2013. "Anaphoric Biases of Null and Overt Subjects in Italian and Spanish: a Cross-linguistic Comparison." *Language and Cognitive Processes.* doi:10.1080/01690965.2013.801502.

Frazier, Lyn. 1999. *On sentence interpretation*. Dordrecht: Kluwer Academic Publishers. doi:10.1007/978-94-011-4599-2

Frederiksen, John R. 1981. "Understanding Anaphora: Rules used by Readers in Assigning Pronominal Referents." *Discourse Processes* 4: 323–347. doi:10.1080/01638538109544525

Garvey, Catherine, and Alfonso Caramazza. 1974. "Implicit Causality in Verbs." *Linguistic Inquiry* 5: 459–464.

Gernsbacher, Morton Ann. 1989. "Mechanisms that Improve Referential Access." *Cognition* 32: 99–156. doi:10.1016/0010-0277(89)90001-2

Gernsbacher, Morton Ann, and David Hargreaves. 1988. "Accessing Sentence Participants: The Advantage of First Mention." *Journal of Memory and Language* 27: 699–717. doi:10.1016/0749-596X(88)90016-2

Gernsbacher, Morton Ann, and David Hargreaves. 1992. "The Privilege of Primacy: Experimental Data and Cognitive Explanations." In *Pragmatics of Word Order Flexibility*, ed. by Doris L. Payne, 83–116. Philadelphia, PA: John Benjamins. doi:10.1075/tsl.22.04ger

Goikoetxea, Edurne, Gema Pascual, and Joana Acha. 2008. "Normative Study of Implicit Causality in Interpersonal Verbs in Spanish." *Behavior Research Methods* 40 (3): 760–772. doi:10.3758/BRM.40.3.760

Gutiérrez-Bravo, Rodrigo. 2008. "La identificación de los tópicos y los focos." *Nueva Revista de Filología Hispánica* 56: 362–401.

Jegerski, Jill, Bill VanPatten, and Gregory D. Keating. 2011. "Cross-linguistic Variation and the Acquisition of Pronominal Reference in L2 Spanish." *Second Language Research* 27 (4): 481–507. doi:10.1177/0267658311406033

Just, Marcel Adam, and Patricia A. Carpenter. 1980. "A Theory of Reading: From Eye Fixations to Comprehension." *Psychological Review* 87 (4): 329–354. doi:10.1037/0033-295X.87.4.329

Keating, Gregory D., Bill VanPatten, and Jill Jegerski. 2011. "Who was Walking on the Beach? Anaphora Resolution in Spanish Heritage Speakers and Adult Second Language Learners." *Studies in Second Language Acquisition* 33 (2): 193–221. doi:10.1017/S0272263110000732

Keating, Gregory D., Jill Jegerski, and Bill VanPatten. 2015. "Online processing of subject pronouns in monolingual and heritage bilingual speakers of Mexican Spanish." *Bilingualism: Language and Cognition* 19 (1): 1–14. doi:10.1017/S1366728914000418

Kras, Tihana. 2008. "Anaphora Resolution in Near-native Italian Grammars: Evidence from Native Speakers of Croatian." In *EUROSLA Yearbook* 8, ed. by Leah Roberts, Florence Myles, and Annabelle David, 107–134. Amsterdam: John Benjamins.

Margaza, Panagiota, and Aurora Bel. 2006. "Null Subjects at the Syntax-Pragmatics Interface: Evidence from Spanish Interlanguage of Greek Speakers." In *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference* (GASLA 2006), ed. by Mary Graham O'Brien, Christine Shea, and John Archibald, 88–97. Somerville, MA: Cascadilla Proceedings Project.

Montrul, Silvina. 2004. *The Acquisition of Spanish: Morphosyntactic Development in Monolingual and Bilingual L1 Acquisition and in Adult L2 Acquisition*. Amsterdam: John Benjamins. doi: 10.1075/lald.37

Phinney, Marianne. 1987. "The Pro-drop Parameter in Second Language Acquisition." In *Parameter Setting*, ed. by Thomas Roeper, and Edwin Williams, 221–238. Dordrecht: Reidel. doi: 10.1007/978-94-009-3727-7_10

Rizzi, Luigi. 1982. *Issues in Italian Syntax*. Dordrecht: Foris. doi: 10.1515/9783110883718

Rizzi, Luigi. 1986. "Null Objects in Italian and the Theory of pro." *Linguistic Inquiry* 17: 501–558.

Rizzi, Luigi. 1997. "A Parametric Approach to Comparative Syntax: Properties of the Pronominal System." In *The New Comparative Syntax*, ed. by Liliane Haegeman, 268–285. New York: Longman.

Rothman, Jason. 2007. "Pragmatic Solutions for Syntactic Problems: Understanding Some L2 Syntactic Errors in Terms of Pragmatic Deficits." In *Romance Languages and Linguistic Theory 2005*, ed. by Sergio Baauw, Frank Dirjkoningen and Manuela Pinto, 299–320. Amsterdam: John Benjamins. doi: 10.1075/cilt.291.19rot

Rothman, Jason. 2009. "Pragmatic Deficits with Syntactic Consequences? L2 Pronominal Subjects and the Syntax-pragmatics Interface." *Journal of Pragmatics* 41: 951–973. doi: 10.1016/j.pragma.2008.07.007

Sagarra, Nuria, and Nick Ellis. 2013. "From Seeing Adverbs to Seeing Morphology. Language Experience and Adult Acquisition of L2 Tense." *Studies in Second Language Acquisition* 35: 261–290. doi: 10.1017/S0272263112000885

Sanford, Anthony J., and Simon C. Garrod. 1981. *Understanding Written Language*. Chichester: John Wiley and Sons.

Sheldon, Amy. 1974. "The Role of Parallel Function in the Acquisition of Relative Clauses in English." *Journal of Verbal Learning and Verbal Behavior* 13: 272–281. doi: 10.1016/S0022-5371(74)80064-2

Slabakova, Roumyana, Paula Kempchinsky, and Jason Rothman. 2012. "Clitic-doubled Left Dislocation and Focus Fronting in L2 Spanish: A Case of Successful Acquisition at the Syntax-discourse Interface." *Second Language Research* 28 (3): 319–343. doi: 10.1177/0267658312447612

Sorace, Antonella. 2011. "Pinning down the Concept of Interface in Bilingualism." *Linguistic Approaches to bilingualism* 1 (1): 1–33. doi: 10.1075/lab.1.1.01sor

Sorace, Antonella and Francesca Filiaci. 2006. "Anaphora Resolution in Near-native Speakers of Italian." *Second Language Research* 22: 339–368. doi: 10.1191/0267658306sr271oa

Stevenson, Rosemary J., Rosalind A. Crawley, and David Kleinman. 1994. "Thematic Roles, Focus and the Representation of Events." *Language and Cognitive Processes* 94: 473–592.

Stevenson, Rosemary J., and Melanie Vitkovitch. 1986. "The Comprehension of Anaphoric Relations." *Language and Speech* 29: 335–360.

Wechsler, David. 1997. *WAIS-III, WMS-III technical manual*. San Antonio, TX: Psychological Corporation.

White, Lydia. 1985. "The "Pro-drop" Parameter in Adult Second Language Learning." *Language Learning* 35: 47–62. doi: 10.1111/j.1467-1770.1985.tb01014.x

Yamashita, Hiroko, Laurie Stowe, and Mineharu Nakayama. 1993. "Processing of Japanese Relative Clause Constructions." In *Japanese/Korean Linguistics: II*, ed. by Patricia Clancy, 248–263. Stanford, CA: Center for the Study of Language and Information.

Index

A

Aboutness Topic 212, 215, 222
Accomplishment 86
Accusative 101, 103–109, 111–112, 114–119, 123, 125, 161
Accusative clitic 101, 103, 105–109, 111–112, 114–116, 123
Achievement 79, 85
Acoustic 192, 197, 206, 230–231, 236, 238–239, 243, 245
Acoustic measures 192
Activity 77–81, 85, 87–90, 93, 95
Activity predicates 78, 80–81, 95
Additive 56–65, 67, 69, 71–75
Additivity 53, 68, 72
Addressee 57–58, 135–136, 138, 140–142, 160, 165
Adjectival clause 34
Age of acquisition 196, 252
Agree 114, 134–135, 185, 212
Agreement 9, 79, 88, 91, 101–103, 105–112, 114, 119–125, 130–132, 136, 152, 165, 171, 186, 188, 209, 211–212, 214, 222, 227
  Markers 108–109, 130–131, 136
  Morpheme 101, 103, 105, 107, 109–111, 114, 121–123
  Morphology 152, 211
ALF 149–150
Algorithm 169, 171, 173, 178–179
Ambiguity 13, 43, 53, 56–62, 68–69, 73–74, 80, 85, 93, 177
American English 191–192, 194–196, 203–204, 207, 229
Anaphor 247–252, 258–261
Anchoring 137
Animate 94, 115, 117–118, 172
Antecedent 105, 218, 247, 249–251, 253–261
Anti-additive 57–60, 62–65, 67, 71–73
Applicative 101, 103, 105, 107–110, 114, 121, 123
  Morpheme 101, 103, 105, 107, 109–110, 114, 121, 123
Aspect 35, 38–39, 41, 48, 79, 81, 91
Aspectual boundedness 78, 95
Aspectual class 79, 82
Assertion 53, 67–68, 74
Ata-nominalisation 92
Ata-nouns 78, 81–85, 87, 89
Atlas linguistique de la France 165
Atomicity 81
Auxiliary 33, 38, 45, 127–131, 133–134, 137–138, 141, 143–145, 218

B

Base-generated 56, 112
Bilingual speakers 191–192, 194–195, 230, 263
Binary feature 159
Breathy voice 192–193, 195

C

Central Southern Italian dialects 127
Change-of-state predicates 79
CHILDES database 216
Class 78–79, 81–82, 129, 135, 242–243
CLD 171–172, 175–176, 178, 180–184
CLDFinder 179–185
CLI 209–211, 213, 215, 218–219, 221–224
Clitic 101–103, 105–112, 114–116, 119–125, 143, 147, 150–152, 160–165, 169, 171–172, 179, 181–188, 226, 264
  see also Dative clitic
  see also Accusative clitic
  see also Direct object clitic doubling
  see also CLI
  see also La
  see also Le
  see also Lo
  Doubling 101–102, 110–111, 115–116, 119–120, 124, 169, 171, 186–188
  Subjects 147
Code errors 180–182
Component 33, 38, 40, 44, 46–47, 49, 72, 91–92, 136, 140, 142
Compositionality of semantics 13, 26
Conjunction 69, 71, 129, 173
Context 31, 56, 58, 60, 62–63, 67, 72–74, 88, 156, 159
Continuous 36, 40–42, 48, 197, 212, 219, 223, 262
Contrast 158–159, 161–163, 165
Convergence 226, 244, 259–260
Copula 36–37, 42–43, 45, 47, 210, 225
Copy 135
Corpora 169, 172, 174–175, 186
Corpus 65, 169–170, 172–177, 186–188, 232, 235, 242, 245
Corrèze 150, 153, 155–156, 158–159
Creaky voice 191–195, 197, 202–204, 206–207
Creuse 150–151, 153, 155–156, 158–159, 235

Cross-linguistic influence 209, 224
CSIDs 127–134, 136, 138–142
Cultural 192, 203–204
Curated text 177, 180, 185

D
Dative 101–111, 114–117, 119–124, 161, 183
  Clitic 101–103, 105, 107–108, 110, 114–115, 161, 183
De Morgan's law 70–71
DE operator 57, 64
Default 15, 26, 28–29, 127, 130, 135–142, 145, 165
Default Marking 127, 130, 138–142
Definite 83, 87, 101, 181
Delimitedness 78–79, 84
Denominal forms 88
Dependency 30, 74, 159, 165, 177, 248–249, 260–261
Derivation 38, 48–49, 81, 85, 89–92, 143
Desire 10, 14, 16–19, 22, 27, 29
Diachronic 82, 145, 147–148, 150, 157, 165, 167, 244
Dialect 101, 115, 128–131, 133, 166, 169, 207, 226, 229, 233
Dialectology 166–167
Diatopic 148, 155, 157, 164
Differentiation 229
Direct Object 108, 109, 169, 171, 179, 186
Direct Object Clitic Doubling 169, 171–179, 183, 186
Directive 10, 14, 16–19, 23, 27, 29
Discourse topics 212, 247–248
Disjunction 70–71
Dislocated 104, 106, 112, 114, 122, 226
Distributed Morphology 89, 91, 127, 130, 143
Distribution 9, 14, 16, 30–31, 34, 44, 53–54, 57–58, 60–63, 67, 69, 72–73, 93, 159, 209–215, 218–219, 222, 230–232, 235, 237, 255
Dominant language 192, 195, 215, 221–222
Dordogne 150, 153, 155–156, 158–159
Double derivation hypothesis 89
Duration 81, 85, 88, 92, 229, 232, 240–244

E
ə 151, 232
ə̀ 229–230, 236, 242–243
Edited text 169, 177
Emotive construal 15, 21
Emotive factives 10
Emotive feature 11, 15, 22, 27–29
Emotive sentience 18, 22
Emotive sentient 18, 27
English 19–20, 45, 50, 53–54, 65, 68, 73, 78, 83, 90, 96, 107, 176, 191–196, 199, 203–207, 209–227, 229–230, 234–235, 242–245, 247–252, 254–258, 260–261, 264
Environment 20, 234
Epistemic 10, 13–14, 16–23, 25, 27–28
Errors 29, 171, 173–175, 178, 180–185, 218, 232, 264
Estar 33–39, 41–52
Even 53–58, 62, 67–68, 73
Event nominals 81
Event plurality 77
Event unity 81
Eventive 33–34, 36, 38, 41–42, 49
Exclusive 65
Expletive 105, 152–153, 155, 159, 218, 226, 248, 250
Extended Projection Principle (EPP) 138, 211–212, 222, 224–226

F
F1 180, 182, 232, 236, 238–239, 241–242
F2 180, 182, 232, 236, 239, 242
F3 232, 236–238, 242
Factive 10, 13–14, 16–19, 21–22, 27–29
Feature geometry 136, 147, 159–161, 164–166
Features 9, 11–12, 15, 17–18, 21, 26–28, 34, 51, 92, 101, 118, 121–122, 127, 130, 132, 134–136, 138–144, 147, 152–153, 159–161, 165, 187, 191, 203, 206, 211, 222, 245–246
Felicitous 40, 56, 58, 63, 65
Felicity 56, 94
Feminine 82, 102, 114–116, 125, 135, 153, 163, 183, 207, 254
Focus 54–55, 57–58, 60, 62–63, 67–68, 72–73
Formant 236, 238
Formation 82, 89, 91–92, 96–97, 226
French 9, 13–14, 16–17, 22–23, 30–31, 53–54, 58–60, 62–63, 65–67, 69, 71–73, 101, 129, 144, 147–148, 150–152, 165, 176, 194–195, 207, 226, 229–235, 237, 240–246, 252
Frenchville 229–231, 233–235, 242–245
Full DP 101, 211
Fundamental frequency 192–194, 196

G
Gallo-Romance 148, 150
Gatineau French 244
Gender 101–103, 114–115, 119, 121–123, 135, 152, 164, 191, 197–200, 207, 254
Gender agreement 101–103, 114, 122
Grammatical person 149, 151–152, 164
Group 135–136, 148, 196–200, 252, 255–258

H
Habitual interpretation 35–37, 39, 41, 48–49
Harmonics 192, 197, 199
Haute-Vienne 150, 153
Heritage speakers 191–192, 195, 199, 209–210, 222, 226, 263
Hexagonal French 232
Homonymy 60, 62

I
IL 34, 49
Imperfective 39–40, 42, 50, 79
Impersonal verbs 158
Impersonal constructions 3, 12, 15, 22–24, 26, 28–29, 157
Inanimate 93, 115, 117–118, 135, 172, 254
Individual-level predicate 34
Individuation 135–136, 139
Infl 134–135, 137–138, 142, 144
Innovation 229–230, 241
Instrumental semelfactive 81
Intrasentential anaphoric subject 247
Island constraints 55, 57, 64
Isogloss 153, 158
Italian 67–68, 77–78, 81, 83, 90–92, 95–97, 101, 127, 129–130, 134, 143–145, 147–148, 151, 176, 213–214, 224, 227, 249–251
Italian dialects 127, 129–130, 143, 145, 147–148, 151

K
Kinds 38, 43, 50, 83, 96, 177
Ko 153, 155–158, 160–165

L
L1 192, 195, 210, 213–214, 220, 222–223, 247–248, 250–253, 258, 260–261
L1 transfer 247–248, 251–253, 258, 260–261
L2 191–192, 195–200, 202, 213–214, 222–223, 247–248, 250–253, 259–261
L2 Acquisition 225, 247–248, 250, 253, 264
L2 learners 191–192, 195, 199, 247–248, 250–251, 259–261
La 101–103, 106–109, 112–113, 115–124, 163, 164, 181, 183, 196, 209
La-for-las 101, 122–123
Laísmo 101, 114–120
Laísta 101–103, 114–116, 118, 120–123
Language background questionnaire 252
Language change 170, 206, 229, 244
Language contact 204, 222–223
Language dominance 203–204, 209, 214, 216, 224
Language isolate 229, 231
Laryngealization 192
Le 101–107, 109, 111, 113–125
Le-for-la 103, 118–121, 123
Le-for-les 101–107, 109, 111, 114–115, 118–123
Left dislocation 112, 119–120, 124, 172
Legacy data 229–231, 234, 236, 242–244
Leísmo 115
leísta 115, 124
Les 101–107, 109–115, 118–125, 149, 156
Lexical DP 121
Lexical entry 12, 16, 27, 67–68, 70
Lexical semantics 9, 11, 18, 94, 96
Lexical specialization 214
Licensing 28, 52, 210–213, 215, 218, 222–224, 247, 251
Lip rounding 230, 238
Lo 115, 171, 176–177, 183–184

M
Majority language 210, 215
Map 15, 147, 242
Markedness 127, 130, 135, 138–142
Masculine 102, 114–115, 151–152, 254
Matrix verbs 56, 254
Meaning 15–22, 26, 53, 65–66, 68–74, 89–92
Même 53–54, 58–65, 67, 71–73
Metaphony 127, 133–134, 141, 143
Meteorological verbs 147, 155, 157–158, 161
Mexican Spanish 191–192, 194–196, 203, 249, 263
Minimal 56, 135–136, 138, 140–142, 212, 222
Minimal Morphological Threshold 212, 222
Modal 22–23, 25, 29, 45, 57–58, 63
Modal voice 192–193, 203
Modern Greek 9, 30
Modern Standard Arabic 252
Monosyllabic 236
Monotone 57–58, 60, 62–63, 67, 72–73
Monovalent 135, 147, 159
Mood 9–11, 13–19, 21–22, 26
Mora 131–132
Moroccan Arabic 247, 250–251, 254, 262
Morpheme 12, 101, 103, 105, 107, 109–111, 114, 121–123, 212
Multiple Grammars hypothesis 222

N
Nanosyntax 11–12, 15, 26, 30–31
Native-like 247, 250–252, 258–261
Natural language processing 170, 174
Ne serait-ce que 53–54, 63–65, 71–73
Near-merger 229–230, 233, 243
Negation 53, 55–57, 59–60, 64, 69–74, 218
NIDs 148, 150–152, 155, 159, 165
NLP 170, 173–175, 178, 185
NODs 148–149, 151–152, 160, 164–165
Nomina vicis 78, 81, 84, 93, 95
Nominal base 84, 88
Nominalisations 77–78, 81–85, 87–88, 90–93, 95
Nominative 147, 161, 166, 227
Non-emotive sentience 18
Non-monotone 57–58, 62–63
Non-veridicality 13
Northern Italian Dialects 130, 147–148, 151
Northern Occitan Dialects 147–148, 150–151, 166
NPI 56–57, 59, 61, 64
NSP 258, 260
Null 104, 120, 247–251, 253–260
Null subject 147–149, 153, 161, 164, 211–214, 218, 220–223, 247–250, 256–257

Null subject pronoun 249–250, 256–257
Number 101–103, 106–107, 111, 121–123, 134–135, 152, 212
Number agreement 101–103, 106–107, 111, 121–123

O
ø 132, 152, 176, 220, 229–233, 235, 241–242, 245
œ 229–233, 235–243, 245–246
Object 94, 105–109, 120, 160, 169, 171, 177, 179, 249, 255–260
Occitan 147–151, 160, 165–167
Occitania 148–149
Offset 236
Only 53–56, 62, 65–69
Onset 206, 223, 236, 252
OSP 259
Overt 29, 106, 109, 127–128, 132–134, 138, 141–142, 209–215, 253–258, 260
Overt subject pronoun 211, 213, 215, 218, 247–249, 251, 256–258
Overt-subject language 209, 211, 214

P
Position of the Antecedent Hypothesis (PAH) 249–252, 258–261
Paradigm 129–130, 134, 150, 161–162, 164–165, 247–248
Parameterization 248
Parsing 169, 175–177, 179, 247–248, 250–251, 258–260
Part of Speech 175
Part of Speech (PoS) tagger 169, 185
Partial agreement 101
Partial paradigms 147, 164
Participant 17, 135–136, 139, 160, 162, 181, 200, 234, 236
Participle 35, 42–43, 45, 47, 82, 96, 129, 143–144
Particles 53–55, 62, 65, 67–69, 71–75
Passive clause 33–35, 39, 42, 48–49
Pattern Identification 169, 175
Peninsular Spanish 194, 204, 249–250
Pennsylvania 166, 229, 243, 245
Perception 17–18, 49, 192, 205–206, 232–233
Perceptual evidence 233
Perceptual match 230
Perfective auxiliaries 127, 130–131, 134, 136, 138–142
Personal constructions 15, 24, 26, 28
Phi-features 161
Phonemic 230–231, 233, 242–243, 245
Phonetic transfer 191
Phonetically distinct 230
Phonetics 205–206, 244–245
Phonology 143–145, 204, 245–246
Pitch range 192
Pluractional activities 88
Plural 83, 85, 93–94, 102, 104, 115, 134–135, 162–163
Polish 73, 143
PoS 169, 174, 176, 178–179, 185
PoS tagger 185
Postverbal 104, 172
PPI 55–57, 60–61
Praat 236, 244
Pragmatic 171–173, 209–215, 218, 220–222, 247–251, 260
Precision 171, 173, 180, 182
Preposition 46, 86, 179–180, 182, 229
Presupposition 53–57, 60–61, 64, 67–68, 74, 90
Presuppositionality 172
Privative 33, 161, 165
Pro 106, 114, 120, 210–211, 248–250
PRO 213
Probe 39, 134–135
Processing 170, 174–175, 185, 210, 213, 247, 249–251, 253, 255, 259
Processive 77–78, 80–81, 85, 87–88, 91, 93, 95
Pro-drop grammar 213–214
Proficiency 196, 216–217, 226, 248, 251–254, 259, 261
Progressive 33, 36–37, 40, 42, 46, 79–80, 147, 152
Pronominal 101, 111, 119–120, 153, 159–161, 210–212, 214, 222, 247, 261
Pronoun 105, 109–111, 116, 136, 151–153, 159, 164–166, 211–212, 218, 227, 249, 253–258, 262
Proto-AGENT 17
Proto-PATIENT 17

Q
Quand même 53–54, 60–62, 64, 71–72
Quantifier 40, 56–57, 183

R
Raddoppiamento Fonosintattico 127, 129, 144–145
Raw text 169, 177–178, 182, 185
Recall 171, 173–174, 179–180, 182
Referential subjects 147
Referring Expression 160
Reflexive 161, 165, 171
Relative amplitude (dB) 197
Relative veridical 14, 16, 18, 21
Rhotic 229–230, 235–237, 241–245
Rhoticization 229–230, 236–238, 241
Rioplatense 169, 172, 175, 178, 180–181, 184–187
Rioplatense Spanish 169, 181, 186
Root 89–91, 93, 155–156, 160–161

S
Scalar 53–54, 56, 60–62, 64–65, 67–69, 71–75, 96
Scalar presupposition 54, 56, 60–61, 64, 68
Schwa 4, 229–233, 235–246
Scope 53–64, 71–75, 248, 250
Se 118, 150, 161, 171, 181, 183–184, 236
Self-paced reading 252–255
Semantics-dependent errors 180, 182
Semel 77, 79–81, 84, 87–88, 91, 93–95

Semelfactive predicates 2, 77–79, 81, 88, 92–93, 95, 97
Sentience 17–18, 22, 27–28, 94
Sentient 17–18, 20–22, 27–29, 94
Sequential bilingual 213, 215, 223
Ser 33–39, 41–52
Seulement 53, 62, 65–67, 69–73
Singular 77–78, 81, 83–85, 93–95, 104, 130, 141–142, 162–163, 165, 181, 254
SL 34, 49
Socio-indexical meaning 203–204
Southern Occitan 148–149
Spanish 19, 33–35, 38–39, 42, 44–45, 49, 51, 67, 101–104, 107, 111, 114–116, 118–119, 121–125, 129–130, 143, 147–148, 165, 169, 171, 174–176, 181, 183–184, 186–188, 191–196, 199, 203–204, 206–207, 209–227, 244, 247–264
Speaker 10, 13–15, 24–26, 28–29, 78, 135–136, 138–142, 152–153, 160, 165, 200
Spectral 192, 229–230, 232, 236, 238–239, 241–243, 246
Stage level predicate 34
State 15, 17–18, 20, 29, 36, 42–49, 79–80, 87, 92
Stative 15–16, 27–28, 33–34, 36, 38, 41–42, 47, 49
Strong veridical 14, 16, 18, 21
Subject 10, 13–15, 17–22, 24, 26–29, 108, 134–135, 147–155, 161, 164–165, 196–197, 209–215, 218–223, 247–252, 255–261
Subject omission 210, 213, 220, 225
Subject pronoun 152–153, 159, 164–166, 227, 249
Subjunctive 9–12, 14–18, 20–23, 25–26, 28–31, 63
Sub-phonemic 230, 242–243
Superset Principle 12
Syllable 131, 197, 231, 235, 254
Syncretism 12–13, 16, 26, 143
Synonymy 60, 62
Syntactic structure 12, 38, 121–122, 138
Syntax-semantics 33

T
Tagging Errors 180, 182
Telicity 45, 78–79, 86, 96
Temporal interpretation 33, 35–36, 39, 41–43, 48–49
Temporoaspectual property 33
Tense 35–36, 39, 42, 48, 127, 136–139, 142, 155–156, 211
Token 82–83, 237, 240–241
Tokenisation effect 82
Topic Criterion 212–213
Topichood 247
Transmission 229, 241, 245
TreeTagger 169, 178, 185
Tweets 170
Twitter 170, 187
Twitter posts 170

U
[*U*coincidence] 137
Underspecified 73, 137, 147, 160–161, 165–166, 213, 220
Upper Southern Italian dialects 127
USIDs 127–129, 131, 142

V
Values 134–135, 138–139, 165, 197–203, 212, 236–239
Verbal base 84, 88–89
Verbal morphology 150, 155, 211–213
Verbs of saying 10, 14, 16–19, 21–22, 27–28
Veridical 9–10, 13–14, 16, 18, 21–22, 28, 30
Viewpoint 35, 39–41, 48, 50
Vocal folds 192–193, 196
Vocal fry 192, 204, 207
Voice quality 191–192, 194–195, 197, 200, 203–207
Vowel 82, 133–134, 206, 229–232, 235–243, 245–246
Vowel backness 239
Vowel height 239, 241–242
Vowel systems 229

W
Web corpus 169–170, 173
Working memory 252–254, 262

This collection brings together current research on a range of phenomena in French, Spanish, Occitan and Italian, that will be of interest to scholars and students of Romance and general linguistics. The volume includes 12 peer-reviewed articles, first presented at the 44th Linguistic Symposium on Romance Languages (LSRL), divided into three sections on syntax-semantics, morphosyntax, and bilingualism and language acquisition.

ISBN 978 90 272 0391 5

John Benjamins Publishing Company